Academic Education Publishing House
Higher Education and Practice Vol. 1 No. 7, 2024

Graduate Employment Quality and College Training Evaluation Based on Machine Learning

Yuling Liu, Can Hou, Shaoyong Hong
School of Artificial Intelligence, Guangzhou Huashang College, Guangzhou, Guangdong, China

Abstract: This study aimed to analyze the influencing factors of talent cultivation quality on the employability of college students. With the increasing demand for applied talents, the education model of colleges and universities is facing challenges. This article collected employment information from 2628 graduates through a questionnaire survey, focusing on the impact of course design, practical training, and teaching on the quality of students' employment. The Lasso model and decision tree analysis method are used to mine and process data. Decision tree analysis helps identify the main factors that affect employability and has strong interpretability. The results indicate that the combination of practical courses and professional knowledge improves students' employability, and soft skills (communication skills) also play a key role in the employment process. According to the results of this study, suggestions for improving the talent-training mode in universities are proposed. To provide empirical support for significantly improving the employment competitiveness of graduates.

Keywords: Employability; Talent Training in Colleges; Lasso Model; Decision Tree; Factor Analysis

1. Introduction

The employment quality of college graduates is a hot issue in society and the focus of public attention[1]. In the context of the new era, cultivating high-quality talents is the main goal of facing social needs and improving employment competitiveness[2]. With the issuance of the 14th Five-Year Plan for Employment Promotion, the goal of achieving higher-quality employment has been further clarified[3]. Institutions of higher education are the main fronts for the cultivation of high-quality talents in China. At the same time, higher education institutions are also a key stage for college students to explore employment directions and cultivate employability[4]. In the information society with the continuous development of big data and artificial intelligence technology, the influencing factors of graduate employment quality and the key path of the employment quality evaluation system strategy of colleges and universities are analyzed by using machine learning methods[5].

Employability is the combination of various abilities possessed by an individual that are conducive to employment[6]. The concept of employability is proposed by the job competition market, describing the collection of skills and abilities, the ability to achieve employment ideals, meet social needs, and realize personal value through continuous learning and comprehensive quality cultivation[7]. Employability is a major factor affecting the quality of employment, and cultivating the employability of college students is a subjective and objective prerequisite for solving employment difficulties and improving employment quality, and is the basis for high-quality employment of college students[8].

Colleges and universities are an indispensable part of education, and the employment quality of college graduates is an important criterion to measure the quality of teaching education and talent training in colleges[9]. Meeting the needs of employers, mastering professional knowledge, and possessing comprehensive employability are the ultimate goals of colleges in cultivating high-quality talents[10,11]. Higher education institutions are addressing the bottleneck problem of improving the skills required for graduates [12,13]. Universities have set up a variety of models in talent training, among which curriculum & teaching systems play an important role[14]. It is crucial to evaluate whether college support the career

needs of graduates through machine learning and intelligent models[15].

2. Data Preparation

2.1 Data Collection

This study mainly collected data in the form of questionnaires, and the target group was graduates of a university in Guangdong Province. The questionnaire design includes five employment quality evaluation indicators: the relevance of work to profession, monthly income, employer size and job position. In addition, the questionnaire also covers the items studied during the university, such as professional courses, public courses, general courses, self-learning ability, activity experience, scientific research projects, competition participation. Respondents were asked to select items from the list that would be helpful to their current job positions, in order to analyze the impact of the university talent training system on the work ability of graduates, and to evaluate the fit between the employment quality of graduates and the talent training model of universities.

A total of 2900 questionnaires were distributed, and after classification and screening, invalid questionnaires were removed, and 2628 valid survey data were collected. These 2628 valid samples will be used to establish a statistical evaluation model for talent cultivation in universities that consistent with the employability of college students. In order to optimize the model, the 2628 samples that could be obtained were divided into two sample sets, which were used for model training and testing. Based on successful recognition experience, selected 1752 samples (approximately 2/3) for training and 876 samples (approximately 1/3) for testing. In order to make the training samples objective and representative, a random selection method is used to divide them.

2.2 Data Analysis and Processing

The type of employer, monthly income (2,000 to 18,000 yuan), size of employer (less than 50 to more than 3,000 employees), job type (marketing, affairs, technology, management), job position (middle management and senior management, ordinary employees), It covers various fields such as industry, education, medical care, and finance, including agriculture & food & chemical industry, heavy industry, economy & finance, religion & entertainment, education & medical care, logistics & communication & information, light industry, and administration & service, a total of eight categories.

Based on the collected data of the respondents, we analyze the majors they studied in college and the occupations they have engaged in after graduation, so as to analyze the matching of majors and employment directions in the employment quality of graduates from the perspective of college training models.

In the questionnaire, four employment quality assessments were designed, namely whether the job is related to occupation, monthly income, employer size, and job position, to measure the employment quality of the respondents. The digitization of these indicators can be carried out as follows:

Whether a job is related to a profession can be represented by a binary variable, i.e., 1 means that the respondent's job is related to their profession and 0 means that it is not relevant.

Monthly income is a metric that can be quantified based on a specific income level. For example, monthly income can be divided into several levels, such as less than 3,000 yuan, 3,000-5,000 yuan, 5,000-8000 yuan, 8000-10 000 yuan, 10000 or more yuan, and give 1, 2, 3, 4, 5 ratings.

For the size of the employer (i.e., the number of people), ranging from less than 50 people to more than 3,000 people, we divide the data into five categories:

Less than 50 people, 51-300 people, 301- 500 people, 501-1000 people, 1001-3000 people, and more than 3,000 people.

Job metrics can be quantified based on the respondent's job level. For example, it can be divided into beginner, intermediate, advanced, management, etc., given a score of 1, 2, 3, and 4 respectively.

By quantifying the above four indicators, a digital employment quality evaluation score can be obtained to measure the employability of respondents. The higher this score, the better the quality of the respondent's employment.

3. Methodology

3.1 Lasso Model Analysis

The Lasso model (Least Absolute Shrinkage

and Selection Operator) is a linear regression method with L1 regularization. The main feature of Lasso is that by introducing L1 regularization term, the regression coefficients of unimportant features can be reduced to zero, thereby achieving feature selection and reducing the complexity of the model. In this analysis, the Lasso regression model was used to evaluate the impact of each dimension on the monthly income of college students and automatically screen for factors that have a significant impact on monthly income.

Lasso regression normalization uses formula (1) objective function to estimate old coefficients:

$$min_{\beta_0,\beta_1,\cdots,\beta_n}\left(\sum_{i=1}^{m}(y_i-\beta_0-\sum_{j=1}^{n}X_{ij}\beta_0)^2+\alpha\sum_{j=1}^{n}|\beta_j|\right) \quad (1)$$

Among them:

y_i represents the amount of revenue (monthly income).

x_{ij} represents self-renewal (different degrees, such as curriculum design, practice, etc.).

β_i represents the regression coefficient of each self decay quantity.

α is the normalization parameter that controls the weight of normalization.

3.2 Results of Lasso Model Analysis

Through the Lasso regression model, the optimal regularization parameter α is selected as 55.902, which achieves the best balance in the model, that is, it not only reduces the complexity of the model, but also retains the features that have an important impact on monthly income.

The important features of Lasso regression screening and their corresponding regression coefficients are shown in Table 1.

Table 1. Important Features and Regression Coefficient Table

serial number	dimension	Lasso coefficient
27	What kind of school experience did you have during your college years (joining the party in school)	363.498586
21	Did you participate in international exchange or study during your university years?	359.039067
15	Did you participate in a research project during your university years?	331.707930
13	The experience of interacting with other classmates	225.532618
25	What kind of school experience did you have during your college years (club leader).	183.850959
3	Public electives	179.524881
9	The general education that has an important impact on career development is (physical education)	158.605713
16	Did you participate in science, technology or major-related competitions (e.g., "Challenge Cup" competitions, mathematical modeling competitions, industrial design competitions) during your university years)	125.817672
23	You did not participate in any of the above activities during your college years	124.437818
6	The general education that has an important impact on career development is (History)	69.746292
24	What kind of school experience did you have during your time at university (student leader)	50.512934
11	General education that has an important impact on career development is (other)	44.421969
5	The general education that has an important impact on career development is (philosophy)	38.929873
26	What kind of school experience did you have during your college years (with awards at the university level and above)	26.177241
2	Public compulsory courses	1.465312
18	Did you participate in club activities during your college years?	0.910170
0	Compulsory courses for majors	-33.652682
22	Did you participate in a social practice activity during your university years?	-82.820723

3.3 Decision Tree Analysis

A prediction model based on decision tree prediction model can be established by investigating whether college students need to improve their professional compulsory courses and professional elective courses, whether they have participated in scientific research projects, major-related competitions, social practices and awards, international exchanges, science and technology. A decision tree model is a simple, easy-to-use nonparametric classifier. It

does not require any prior assumptions about the data, is computationally fast, the results are easy to interpret, and robust[16] Therefore, in the problem of predicting the employability of college students, which has a multivariate relationship and needs to consider multiple factors, the decision tree regression model can better reflect the complex correlation between the variables, so as to make a more accurate prediction of the target variables.

The decision tree method is used to identify the key factors that affect the employability of college students. Establish a decision tree model using the CART algorithm, with employability as the target variable and feature variables as influencing factors. By constructing a decision tree, analyze and visualize the importance of different factors in the decision-making process. In order to determine which dimension of each node in the decision tree is divided on, and to determine on which value the features of the selected dimension are divided, we introduce the Gini index based on the Gini coefficient, which reflects the probability that two samples are randomly selected in the dataset and their category flags are inconsistent, and formula (2) is as follows:

$$D(p) = \sum_{k=1}^{k} p_k(1 - p_k) = \sum_{k=1}^{k} p_k^2 \quad (2)$$

The more impure the node and the larger the GINI value, the worse the node classification or prediction effect, and conversely, the smaller the Gini index, the higher the purity after partitioning. Decision tree regression models are suitable for predicting the values of continuous variables. In the prediction of college students' employability, This study constructed a decision tree model based on the following characteristics: investigating whether college students need to improve their required and elective courses in their major, major-related competitions, whether they have participated in international exchanges, social practice and awards, etc., a decision tree model can be constructed by learning the existing employment data of college students, and then the model can be used to predict the employability of new college students. In the problem of predicting the employability of college students, which involves multivariate relationships and requires comprehensive consideration of multiple factors, the decision tree regression model can better reflect the complex correlation between various variables, so as to make more accurate predictions of the target variables.

A large number of feedback data from college students were collected, including whether college students' major compulsory courses and major elective courses needed to be improved, whether they had participated in research projects, social practice experience and awards, major-related competitions, and other factors, so that to Build a dataset containing multiple features. In addition, as much as possible, we will collect as much feedback data as possible, and clean, deduplicate, and fill in the missing values of the data. The eigenvalues in the read dataset are split into X sets, and the label values are split into Y sets. The dataset is randomly divided into a training set, a test set, and a validation set. When building the model, 60% of the data is used to train the model, 20% of the data is used to test the model, and the remaining 20% is used for the validation set. Build and train a model using DecisionTreeRegressor() to create a decision tree regression model, and then train it using the fit() method. Use the predict() method to predict the Y value corresponding to the X value in the test set using the model.

Table 2. Model Parameter Table

Name	Parameter name	Parameter values
Model parameter setting	data preprocessing	None
	Training Set Proportion	0.6
	node splitting criterion	gini
	Node division method	best
	Node split samples	25
	Number of terminal nodes	13
	Maximum tree depth	5

In order to achieve better prediction effect, the parameters are adjusted and optimized according to the specific data set and problems, the relevant parameters of the decision tree model are detailed in Table 2, the proportion of the training set is 60%, the node splitting criterion is the splitting criterion of the decision tree node is the Gini impurity, the node division method is the division of the decision tree node is the binary tree, and the number of node splitting samples is the minimum number of samples of the decision tree node splitting is 25, which is used to control the complexity and generalization

ability of the decision tree. The number of terminal nodes is 13, which is used to control the complexity and generalization ability of the decision tree, and the maximum depth of the tree is the maximum depth of the decision tree is 5, which is used to control the complexity and generalization ability of the decision tree.

The mean square error (MSE) and coefficient of determination (R-squared) were used to evaluate the performance of the decision tree regression model to predict the employability of college students.

1. Mean Square Error (MSE): Represents the average of the square values of the error between the actual value and the predicted value. The smaller the MSE, the better the prediction performance of the model.

where denotes the sample size, denotes the actual value, and denotes the predicted value. Using the Mean Square Error (MSE) to evaluate the performance of a decision tree regression model that predicts the employability of college students can be achieved with the following code: where y_test is the true value on the test set and predict is the result of the prediction. We want the MSE to be as small as possible, which means that the difference between the value predicted by the model and the true value is smaller and closer to the actual value. The specific unit of the MSE is the square of the y-value, so it depends on the size of the actual value.

2. Coefficient of Determination (R-squared): Represents the linear correlation between the predicted value and the true value, i.e., the percentage of the variance that the model is able to explain for the data. red can range from 0 to 1, with closer to 1 than the model fitting better. The formula for calculating R-squared is: Using the coefficient of determination (R-squared) to evaluate the performance of the decision tree regression model that predicts the employability of college students can be achieved with the following code: where y_test is the true value on the test set and predict is the result of the prediction. We want the value of R-squared to be as close to 1 as possible, which means that our model is able to account for most of the changes, i.e., accurately predict the target variable. Evaluating the decision tree regression model with the coefficient of determination (R-squared) and the mean square error (MSE) together provides a more complete understanding of the model's performance.

The final result is that the MSE of the training set and the test set are very small, which indicates that the model fits well and can predict the employability of college students well. Moreover, the R-squared of the training set and the test set are close to 1, the model has strong explanatory properties, which can well explain the changes in college students' employability.

4. Conclusion

Through the analysis of Layso and decision tree, this study deeply explores the impact of the quality of talent training on the employability of college students. The results indicate a significant positive correlation between curriculum design and employability, especially practical courses and professional courses, which can effectively improve students' professional skills and help them better adapt to the needs of the workplace. This finding suggests that colleges should pay more attention to the connection with market demand in curriculum design.

In addition, the findings highlight the importance of hands-on training. Students who participate in internships and research projects demonstrate excellent employability, indicating that their rich practical experience not only enhances their professional level, but also improves their ability to solve real-world problems. The results of decision tree analysis further reveal that the impact of practical experience on employability is significantly different in different situations, which provides a reference for the optimization of talent training programs in colleges and universities

Personal qualities are also identified as important factors that affect employability. Students who have strong self-learning abilities and actively participate in extracurricular activities are usually more competitive in the job market. This is consistent with the conclusion of the existing literature, emphasizing the importance of comprehensive quality education, and suggesting that colleges and universities should pay more attention to cultivating students' soft skills in the process of talent training to improve their overall employability.

Finally, the survey results show that monthly

income and job title significantly affect the employment satisfaction of graduates, reflecting that higher salaries and higher positions are often closely related to better educational background and practical experience. These findings provide an empirical basis for the education reform of colleges and universities, and it is suggested that the cooperation with employers should be strengthened in the design of courses and practices, so as to cultivate applied talents that are more in line with market demand.

In summary, this study reveals the multi-faceted impact of the quality of talent training on the employability of college students, and provides an important reference for further improving the talent training model of colleges and universities.

References

[1] Yang Heqing, Li Jia. An empirical analysis of the employment quality of college graduates. China Labor, 2007(12): 26-28.

[2] Zhang Shaojie. Cultivating Foreign Language Talents with "Solid Foundation, Strong Ability and High Quality" for the Needs of Diverse Society and Multiple Goal Orientations: New Thinking on the Teaching Reform of English Major Education. Chinese Foreign Languages, 2010, 7(03): 4-9.

[3] He Yuanya. A Study on the Comprehensive Quantitative Evaluation of the Employment Quality of College Graduates: A Case Study of T College. Journal of Higher Education, 2023, 9(27): 106-108+112.

[4] Zhu Xin Scale. On the Cultivation of College Students' Employability. Higher Education Exploration, 2009(04): 124-127.

[5] Wang Wenzhong. Current situation and countermeasures of college students' employment under artificial intelligence. Employment of Chinese College Students, 2018(02): 10-12.

[6] Wang Yi.: an important aspect to promote the employment of college graduates. Educational Development Research, 2005(07): 31-34.

[7] Zheng Xiaoming. Employability". Journal of China Youth University of Political Science, 2002(03): 91-92.

[8] Qin Jianguo. Analysis on the evaluation system of college students' employment quality. Reform and Strategy, 2007(01): 150-153.

[9] Cao Yuxia. Analysis of influencing factors and countermeasures of employment of graduates of private college graduates. Education and Career, 2017(17): 65-69.

[10] Yan Dawei. On the composition and elements of college students' employability. Qinghai Social Sciences, 2007(06): 28-31.

[11] Wu Zhongjiang, Huang Chengliang. The connotation of applied talents and the cultivation of applied undergraduate talents. Research on Higher Engineering Education, 2014(02): 66-70.

[12] Y. C. Gu and C. S. Zhao, “Innovative research on the cultivation system of the core competence of college students,” Educ. Sci. Pract., vol. 18, no. 6, pp. 3844–3851, 2018.

[13] B. Mulvey, “International Higher Education and Public Diplomacy: A Case Study of Ugandan Graduates from Chinese Universities,” High. Educ. Policy, vol. 33, no. 3, pp. 459–477, 2020.

[14] Dong Zefang.Definition of the concept and analysis of the elements of the talent training model in colleges and universities. University Education Science,2012(03):30-36.

[15] M. F. Bashir, H. Arshad, A. R. Javed, N. Kryvinska, and S. S. Band, “Subjective Answers Evaluation Using Machine Learning and Natural Language Processing,” IEEE Access, vol. 9, pp. 158972–158983, 2021.

[16]Korea Miao.Research on Graduate Employment Prediction Based on Decision Tree Algorithm. Computer Programming Skills and Maintenance,2020(4):64-67.