Salary Prediction Analysis for the 'Slow Employment' Phenomenon - Based on Random Forest Algorithm

Liang Ye[1], SiYi Fu[2], GuangYuan Chen[2], and Jin Lu[1](✉)

[1] Office of Admissions and Career Services, Shanghai University, Shanghai, China
liangye@shu.edu.cn, lujin221@t.shu.edu.cn

[2] School of Computer Engineering and Science, Shanghai University, Shanghai, China

Abstract. The "slow employment" issue among college students is a contemporary employment problem that requires attention. To address this issue and enhance students' employability, this study uses a university in a developed region as a case study and analyzes data, conducts a questionnaire survey, and performs literature analysis to categorize "slow employment" into active and passive types based on the psychological conditions of economically developed college students. The results show that the highest percentage of slow employment (74.08%) among students in general undergraduate colleges and universities is due to the training methods of schools and the lack of students' career awareness. The study also examines job skills and salary forecasts to identify the root causes of slow employment among students and proposes corresponding countermeasures.

Keywords: Slow employment · Random Forest · Grid search method · Salary prediction

1 Introduction

As higher education becomes more widespread, the issue of "slow employment" among recent graduates has become increasingly apparent. According to the "2022 China College Student Employment Report," released by McKeith Research Institute, 17.46% of graduates from the class of 2021 are not employed, and slow employment negatively impacts economic development and social stability [1, 2]. This paper proposes a salary prediction model based on the random forest algorithm to help companies and job seekers understand market value and enable graduating students to better connect with society. After comparing different algorithms, the proposed model was found to have good effects and can address the issue of "slow employment [3, 4]."

2 The Current Situation and Causes of "Slow Employment"

2.1 Current Situation

"Slow employment" is when college graduates delay entering the job market by taking a break, traveling, teaching, staying with their parents, or exploring business opportunities. It can be active, where individuals enrich themselves or start their businesses, or

P. Qi and Z. Chen (Eds.): ICBDIE 2023, AISR 178, pp. 298–303, 2024.
https://doi.org/10.2991/978-94-6463-238-5_40

Table 1. Research data on "slow employment" among graduates from 61 universities in a certain economically developed province (city) before the outbreak of the pandemic.

University type	Slow Employment Type						
	I	II	III	IV	V	total	Ratio
Top university	450	50	10	47	4	561	6.55%
First-class discipline	1063	272	38	144	10	1527	19.37%
Ordinary university	3932	1285	209	387	26	5839	74.08%
total	5445	1607	257	578	40	7882	100%

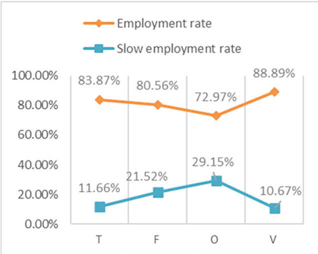

Fig. 1. Employment Rate of Different Educational Levels and "Slow Employment Rate"

passive, due to the competitive job market or lack of opportunities. In a study of 61 colleges and universities, 69.5% of regular undergraduate college and university graduates experienced "slow employment," pursuing further education or civil service exams [5]. Table 1 and Fig. 1 show "slow employment" rates and employment rates for different academic levels.

2.2 Analysis of the Causes of the "Slow Employment" Phenomenon

This study examined the reasons behind the "slow employment" trend among 3,640 students in an economically developed province/city, with 57% undergraduates, 41% master's students, and 2% doctoral students [6]. The survey found that 36% had a positive attitude, 56% were neutral, and 8% believed it was an excuse to avoid reality. The reasons identified were an inability to handle hardship and overconfidence in academic institutions and professional abilities. To address these issues, this paper proposes conducting comparative experiments to predict salary and analyze vocational skills' influence on it. Targeted employment-oriented learning can also be provided to help students enhance their employability and find suitable jobs promptly, addressing the "slow employment" problem.

Table 2. Eigenvalue

Benefits	Five social insurance and one fund, paid annual leave, holiday benefits, total attendance award, meal allowance, regular physical examination, overtime allowance, employee travel
recruitment requirements	The machine learning algorithm, HTML, Java, distributed technology, Vue, Performance testing, Linux, SQL, Python, Android, embedded software development, Embedded software technology, Spring, deep learning algorithm, back end, full stack, front end, Mybaits, Microservices

3 Salary Prediction Experiment

3.1 Eigenvalue Selection

This experiment aims to construct a prediction model for software development job salaries by analyzing the impact of various factors. Feature selection is done by analyzing recruitment requirements through word frequency statistics and manual extraction. The selected feature values are listed in Table 2.

3.2 Contrast Experiment

Cross-Validation Comparison. After cleaning the data, it was divided into two parts: salary data as labels and filtered data as feature values. The data was then split into training and test sets. Four prediction models were constructed using the sklearn module: linear regression, decision tree, KNN, and random forest. The models were trained using the training set and evaluated using the test set, with the mean square error used as the evaluation metric. The cross-validation scores of each model are presented in Fig. 2.

Figure 2 shows that the KNN model's mean score is 6.9825, the random forest model's is 5.9647, the decision tree model's is 7.6020, and the linear regression model's is 5.9943 after 10-fold cross-validation score calculation. The random forest and linear regression models have better prediction results than the decision tree and KNN models, suggesting they are more suitable for this study.

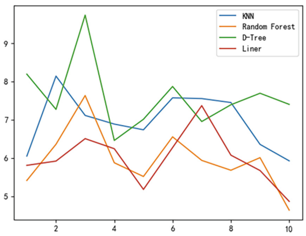

Fig. 2. Cross-validation results of four types of models

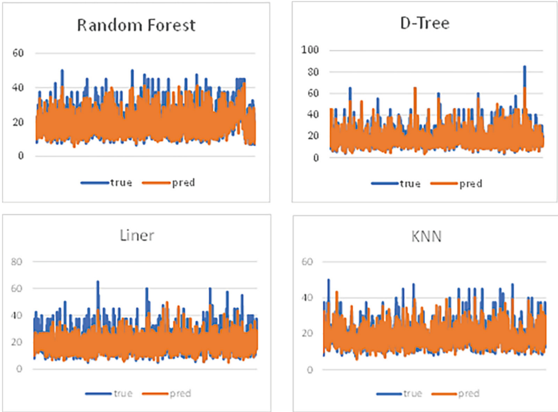

Fig. 3. Comparison of regression results of four types of models

Prediction Accuracy Comparison. In Fig. 3, the highest accuracy is achieved by the random forest model with an accuracy of 0.85368, followed by the KNN model with an accuracy of 0.85210, the decision tree model with an accuracy of 0.85154, and the linear regression model with an accuracy of 0.81856. The comparison of the accuracy of these models shows that the random forest model is the most suitable for this study, despite the 10-fold cross-validation score suggesting that both random forest and linear regression are suitable.

3.3 Analysis of the Influence of Eigenvalues on Wages Based on Random Forest

The eigenvalues represent hiring needs and their impact on salary, which is shown in Fig. 4 and Fig. 5. Figure 4 illustrates the impact of each feature on salary before model training, and Fig. 5 shows the influence of each feature on a salary after training the random forest model. Both figures suggest that the impact of each feature on salary increased after training the random forest model. The results indicate that deep learning skills have a greater impact on salary, which is consistent with the project's focus on software development jobs.

Based on the analysis of the second part of the questionnaire, companies value professional ability more in the job market. The reason for slow employment in ordinary undergraduate colleges and universities is due to the overall teaching system being inferior to first-class universities, and students not recognizing the importance of professional knowledge. The paper will provide countermeasures to address slow employment.

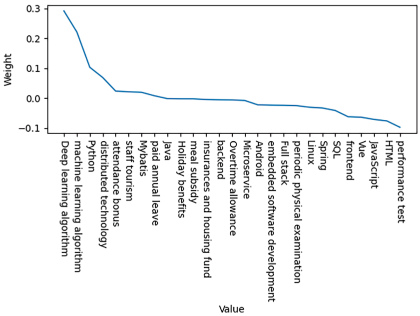

Fig. 4. The influence of eigenvalue before model training

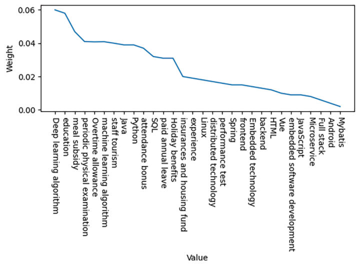

Fig. 5. The influence of eigenvalue after model training

4 Conclusion and Countermeasure

This paper investigated the "slow employment" issue for students from general undergraduate colleges and universities. A questionnaire survey was conducted to analyze the causes of slow employment, and salary prediction experiments were performed to identify the impact of job skills on salary. Based on the findings, the paper proposed several countermeasures for both students and colleges/universities. Enhancing their practical skills and clarifying their future development direction is recommended for students. Colleges should focus on training students' professional abilities, career planning education, and job market knowledge. These measures can help improve employment prospects and reduce the occurrence of slow employment.

References

1. Chen Liming, Liu Binghui, 2019. The Renaissance Generation: An Analysis of the Times Characteristics of "Post-95" College Students [J]. Journal of East China University of Science and Technology (Social Science Edition), 34(04):107-116. (in Chinese)
2. Diao Ying, Xu Baogui, 2020. Analysis and Countermeasure Research of Post-95 Graduates' Career Values [J]. Research on Ideological and Political Education, 36(01):156-160. (in Chinese)
3. Jiang Liping, 2020. Interpretation and governance of college students' "slow employment" phenomenon from the perspective of the transformation of major social contradictions [J]. Contemporary Youth Studies, (06):70-76. (in Chinese)

4. Jiang Liping, Liu Yuwen, 2020. Essence Analysis and Countermeasures of College Students’ “Slow Employment” Phenomenon [J]. School Party Building and Ideological Education, (04):64-66. (in Chinese)
5. Liu Yiying, Xu Ruixue, Li Dejing, 2019. Research on the Countermeasures of “Slow Employment” of College Graduates from the Perspective of Whole-process Employment Guidance [J]. Employment of Chinese College Students, (24):54-59. (in Chinese)
6. Li Ning, 2019. Reasons and countermeasures for “slow employment” of college graduates [J]. People’s Forum, (14):116-117. (in Chinese)