Match Model of Data-Analysis Competencies in Higher Education Institutes and Enterprises

Xiaoyu Wei[1,3,b], Guoqiang Chen[1,2,a], Xichao Zhao[3,4,c*], Feng Yu[3,d], Bowen Chen[3,e], Di Zhang[3,f], Ngan Iek Hang[3,g]

[1]Beijing Institute of Technology, Zhuhai, China
[2]Zhuhai Science and Technology Development Promotion Association, Zhuhai, China
[3]Faculty of Humanities and Social Sciences, Macao Polytechnic University, Macao, China
[4]Guangxi College of Finance and Economics, Nanning, Guangxi, China

[a]Lxchen08@sina.cn, [b]P2209532@mpu.edu.mo,
[c*]Corresponding author: P2209677@mpu.edu.mo, [d]Legalfish@qq.com,
[e]p2312561@mpu.edu.mo, [f]p2109128@mpu.edu.mo, [g]21083505@qq.com

Abstract. The demand for highly skilled technical data analysts has significantly increased due to the growing emphasis on advanced technology and its widespread use in related industries. This study uses quantitative analysis to develop a matching model for data analysis competencies. The aim is to analyze the coordination between curriculum and job requirements in higher education institutions. Conclusions: The data analysis industry has promising prospects, with a rapidly growing demand for jobs. The skill requirements for data analysis align with the educational content provided to practitioners in colleges and universities. Nevertheless, there exists a substantial need for individuals possessing proficient skills in data analysis. The skills necessary for data analysis roles are in line with the skills imparted to professionals in colleges and universities. Nevertheless, there is still a discernible inadequacy in the capabilities of data analysis. Suggestions: training of advanced data analysis skills, specifically in the utilization of data analysis software, should be emphasized in higher education. The cultivation of data analysis proficiency is contingent upon the synergistic endeavors of academic institutions and social enterprises.

Keywords: Data Analysis skills; Match model； Higher education; Coupling Coordination Analysis.

1 Introduction

In 2019, the government disseminated information regarding the introduction of 13 novel professions to the public. These occupations encompassed various fields, such as artificial intelligence engineering and technology, Internet of Things engineering and technology, big data engineering and technology, and cloud computing engineering and technology[1]. In recent years, China has witnessed significant advancements and

Y. Kuang et al. (eds.), *Proceedings of the 2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM 2024)*, Atlantis Highlights in Computer Sciences 22,
https://doi.org/10.2991/978-94-6463-502-7_3

innovations in artificial intelligence, Internet technology, big data, and cloud computing. Consequently, these industries have emerged as key drivers of China's economic growth. Various enterprises have exhibited a favorable response and have progressively embraced the concept of digital transformation.

The demand for data analysis positions is experiencing a significant growth. According to statistical data, it has been observed that a significant majority of foreign enterprises, approximately 90%, possess dedicated data analysis teams. Currently, there is a significant talent gap in data analysis in China, estimated at 1.5 million individuals. This shortage of skilled professionals has resulted in a substantial number of unfilled job positions. Data analysis positions require individuals to possess certain capabilities. These include a strong familiarity with industry knowledge, a mastery of the fundamental principles and effective methods of data analysis, proficiency in commonly used data analysis tools, and the ability to present analytical viewpoints clearly and effectively through the use of charts and graphs[2].

Higher education institutions, being the primary institutions for nurturing data analysis talents, should incorporate the development concept of the new era and integrate the characteristics of the current times and environmental changes into their talent cultivation approach[3]. This will facilitate the improvement of students' knowledge and professional skills, as well as enhance the transformative education of data analysis talents[4]. In recent years, there has been a rapid advancement in the data analysis industry. Further field research is necessary to determine whether the development of the industry has an impact on the data analysis skills of students trained in colleges and universities, as well as whether these skills align with the requirements of corporate positions[5].

2 Methodology

This paper aims to examine the present state of data analysis proficiency and the demand for data analysis jobs in the higher education sector. The present study examines the influence of data analysis proficiency on job market demand. Based on the explanation and justification of the certificate system concept, a survey questionnaire was designed from three dimensions: cognition, ability, and attitude. The questionnaire design consists of two parts. The first part is personal basic information, including gender, age, and the school attended. The second part focuses on the investigation of the certificate system, asking questions about students' cognitive level, attitude, acquisition of vocational skill level certificates, and the courses they have taken. Following the procedures of statistical investigation, the survey was conducted in higher education institutes. Make suggestions based on the statistical results of the questionnaire. The research study employed questionnaires and utilized big data analysis methods for data collection. Statistical analyses were performed using SPSS 20.0 for Windows (SPSS Inc., Chicago, Chicago, Ill., USA).

3 Results

3.1 Samples

Sample Capacity Calculation:

$$N = Z^2 \times \frac{P \times (1-P)}{E^2} \tag{1}$$

A total of 406 questionnaires were distributed to university students, and all 406 were successfully returned. Among the participants, 34.6% were male and 65.4% were female. Additionally, 23.9% of the participants identified as freshmen, 26.7% as sophomores, 21.5% as juniors, and 27.9% as seniors.

3.2 Matching Model

Combining the expertise in data analysis and the current market demand, this study aims to develop a comprehensive model that assesses the competency level required for data analysis and the corresponding job demand[6]. Subsequently, the study systematically evaluates the degree of alignment between the competency level and job demand in the field of data analysis (Figure 1).

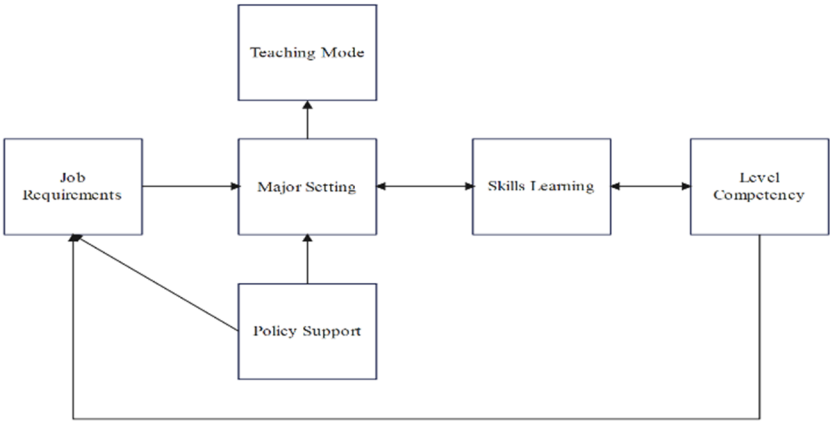

Fig. 1. Match Model

3.3 Compatibility

Introduction to the Model

The series-based matching method requires the analysis and calculation of the matching degree between variables X and Y. The values of variables X and Y in each cell are denoted as (X1, Y1), (X2, Y2), ..., (Xk, Yk). The matching degree of X and Y in different cells is calculated as follows: The values of X for k cells X1, X2, ..., Xk can

be sorted in ascending order, resulting in the corresponding numbers n1, n2, ..., nK. Similarly, the Y values Y1, Y2, ..., and YK of the k cells can be arranged in ascending order, and the corresponding numbers are denoted as m1, m2, ..., and mK...., mK. When X and Y have larger values, the two variables exhibit a higher degree of correlation. The degree of compatibility, denoted as π, is calculated using the compatibility formula as follows:

$$\pi_i = 1 - \frac{|n_i - m_i|}{K-1} (i = 1,2, \dots, K) \tag{2}$$

The level of compatibility can be quantified by calculating the disparity between the variables using the formula (2). When ni is equal to mi, it indicates a perfect match, resulting in a degree of compatibility π equal to 1. Conversely, as the difference between ni and mi increases, the compatibility decreases, and the degree of compatibility approaches zero.

When X has a higher value and Y has a lower value, the two variables exhibit a greater degree of compatibility. The matching degree π is calculated using the compatibility formula as follows:

$$\pi_{\mathrm{i}} = 1\text{-}\frac{|\mathrm{n_i\text{-}m_i\text{-}K\text{-}1}|}{\mathrm{K\text{-}1}} (\mathrm{i} = 1,2, \dots, \mathrm{K}) \tag{3}$$

According to Equation (3), a larger difference between ni and mi results in a matching degree πi that is closer to 1, while a smaller difference between ni and mi leads to a matching degree πi that is closer to 0.

Alternatively, the match can be determined by calculating the ratio of the specific value of the variable in each cell to the total value of that variable across the entire study area.

There are two distinct scenarios, each with its corresponding formulas for calculation.

$$\pi_i = 1 - \frac{|r_i - s_i|}{max(r_k, s_k) - min(r_k, s_k)} \tag{4}$$

$$\pi_i = 1 - \frac{|r_i + s_i - max(r_k, s_k) - min(r_k, s_k)|}{max(r_k, s_k) - min(r_k, s_k)} \tag{5}$$

$$\mathrm{r_k} = \frac{\mathrm{x_k}}{\sum_{\mathrm{i=1}}^{\mathrm{k}} \mathrm{x_i}} \mathrm{s_k} = \frac{\mathrm{y_k}}{\sum_{\mathrm{i=1}}^{\mathrm{k}} \mathrm{y_i}} (\mathrm{k} = 1,2, \dots, \mathrm{K}) \tag{6}$$

Data Sources

The present study employed big data analytics methods to analyze skill requirement information in the data analysis industry, specifically focusing on data obtained from Wisdom Link Recruitment and MileagePlus. Upon analyzing the skill requirements, it was found that the four most frequently occurring skill keywords are SQL (27%), Python (17%), SPSS (15%), and Excel (12%) (Figure 2).

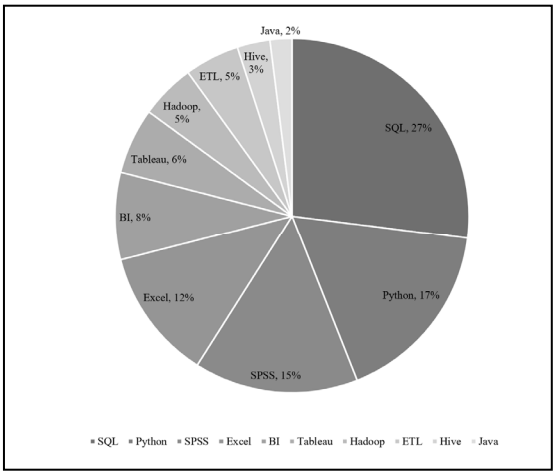

Fig. 2. Percentage of data analysis software requirements

The findings of the survey conducted on the data analysis courses provided by the university are presented in Figure 3. The four most commonly offered courses, in descending order, are Excel (24%), Python (19%), SQL (13%), and SPSS (10%).

Compatibility Analysis

The statistical analysis of the aforementioned survey indicates that the skill requirements and data analysis course curriculum remain consistent, with the top four skills consistently identified. Consequently, the selection of SQL, Python, SPSS, and Excel was made to conduct the compatibility analysis in this study.

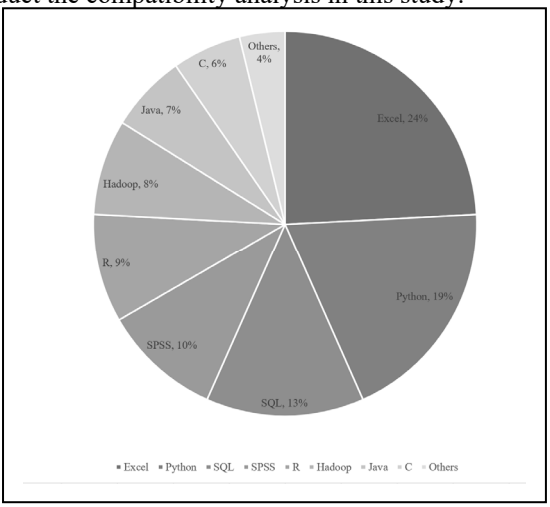

Fig. 3. Percentage of data analytics software curricula in higher education institute

The concept of a match can be operationally defined as a match when the degree of similarity reaches or exceeds 0.8. It can be considered a better match when the similarity falls within the range of 0.8 to 0.6. Conversely, a mismatch is identified when the similarity is less than or equal to 0.6. Based on this criterion, the analysis and calculation of the match between the different variables is conducted.

As indicated in Table 1, there is a strong correlation between Python and SPSS, suggesting that the curriculum for these courses aligns well with the requirements of data analysis roles. SQL and Excel exhibit a significant disparity in terms of their compatibility, giving rise to two distinct factors contributing to this mismatch. Firstly, there is a high demand for skills in data analysis positions, yet the curriculum in higher education is relatively limited in terms of topics such as SQL. Secondly, there is a lower demand for job skills, although the curriculum is more prevalent in universities. The enrollment for the Excel course has exceeded capacity. The aforementioned factors contribute to a discrepancy between the curriculum and the requisite job skills for data analysis.

Table 1. Degree of Compatibility between Skill Requirement and Courses Curriculum

	SQL	***Python***	***SPSS***	***Excel***
Skill Requirement	27%	17%	15%	12%
Courses Curriculum	13%	19%	10%	24%
Sequences in Skill Requirement （From small to large）	4	3	2	1
Sequences in Courses Curriculum （From small to large）	2	3	1	4
Degree of Compatibility	0.33	1	0.66	0.12

Curriculum and Job Requirements Coupling Coordination Analysis

The degree of coupling coordination, denoted as D, is conceptualized as the capacity coupling coefficient in the field of physics. The coupling degree, denoted as C, and the coordination index, represented by T (Formula 7), are both important measures in this context. The classification criteria for the coupling coordination degree class are presented in Table 2. The findings pertaining to the degree of coupling coordination between the curriculum and job requirements are presented in Table 3.

$$D = \sqrt{C \times T} \tag{7}$$

Table 2. Criteria for Evaluating the degree of coupling coordination

Interval of Coupling Coordination D	***Level***	***Degree of coupling coordination***
(0.0~0.1)	1	Extremely Disordered
[0.1~0.2)	2	Severely Disordered
[0.2~0.3)	3	Moderately Disordered

[0.3~0.4)	4	Mildly Disordered
[0.4~0.5)	5	Nearly Disordered
[0.5~0.6)	6	Barely Coordinated
[0.6~0.7)	7	Elementarily Coordinated
[0.7~0.8)	8	Intermediately Coordinated
[0.8~0.9)	9	Well Coordinated
[0.9~1.0)	10	Perfectly Coordinated

Table 3. The degree of coupling coordination

Item	***C Value***	***T Value***	***D Value***	***Level***	***Degree of coupling co-ordination***
SPSS	0.884	0.452	0.632	7	Elementarily Coordinated
SQL	0.245	0.327	0.283	3	Moderately Disordered
Excel	0.335	0.173	0.241	3	Moderately Disordered
Py-thon	1.000	0.990	0.995	10	Perfectly Co-ordinated

4 Conclusions

1. The significance of data analytics capabilities is expected to grow significantly in the future. The analysis of the demand for data analysis skills across different industries reveals a consistent and significant increase in the demand for individuals with expertise in data analysis. This upward trend in demand for data analysis talents has been observed over the years. With the increasing number of enterprises embracing digital transformation, coupled with the maturation of information technology infrastructure, the market demand for positions related to data analysis has been significantly stimulated. Between 2011 and 2014, China's big data industry was in its nascent stage, experiencing an annual growth rate exceeding 20%. In 2015, the market size of big data reached 9.89 billion yuan. The growth rate in 2016 reached 45%, amounting to over 16 billion yuan. In the year 2020, it is projected that the market size of big data in China will surpass 800 billion yuan.

2. The demand for data analysis skills and the availability of skills-learning courses are generally aligned, although there is still a need to ensure that the curricula are in line with industry demands and to increase the number of courses accordingly. The findings of the survey indicate that respondents continue to exhibit a significant deficiency in their utilization of advanced data analysis skills. This suggests that their proficiency is limited to the mere completion of learning tasks[7]. According to the findings regarding the matching degree and coupling degree of coordination between the curriculum and job demand, it has been observed that the Python and SPSS curricula are more aligned with the skill requirements of the job market. However, it is worth noting that the curriculum of higher education does not fully align with the skill demand[8]. It is

advisable for higher education institutions to enhance the curriculum by incorporating additional courses on Python and SPSS, while reducing the emphasis on Excel courses.

3. The research findings indicate that the participants expressed a strong desire for acquiring knowledge in data analysis for professional learning. They also expressed the hope that the school would offer relevant courses and strengthen collaboration with enterprises to enhance their learning experience. The survey findings indicate that a majority of the participants express a strong inclination towards specializing in data analysis. Additionally, they express a desire for the acquired knowledge in data analysis to be applicable in their job search post-graduation. Furthermore, the respondents exhibit interest in pursuing courses and training opportunities that encompass both on-campus and off-campus settings. The level of acceptance towards the implementation of courses that integrate both on-campus and off-campus training is substantial.

5 Suggestions

1. It is crucial to prioritize the development of data analysis skills by implementing relevant courses in educational institutions and actively engaging in internal and external collaborations[9]. This approach aims to foster a greater number of proficient data analysts and enhance the overall capacity of data analysis education in higher education institutions.

2. Strengthening the training of data analysis ability, particularly the curriculum system pertaining to data analysis software, is of utmost importance. In response to the demands of the data analysis industry, we are developing the curriculum for vocational skills level certificates that focus on specialized skills, with the aim of enhancing students' knowledge in this field. The justification for the selection of each specialized course should be thoroughly examined in order to emphasize the unique characteristics of the specialty. Additionally, it is crucial to incorporate industry skill requirements, which can be demonstrated through the attainment of vocational skills level certificates. In the endeavor to cultivate data analysis talents, it is imperative to not only prioritize the acquisition of general-purpose analysis software skills, but also to align with the specific skill requirements of data analysis roles[10]. Proficiency in Python programming and SQL database management is essential for effective data analysis.

3. The development of data analysis skills relies on the collaborative efforts of higher education institutions and social enterprises. Colleges and universities play a crucial role in nurturing data analytics talents, which serve as a valuable resource for enterprises. In order to enhance this partnership, it is imperative for educational institutions to strengthen their collaboration with enterprises. The demand for data analysis in the industry closely aligns with and provides insight into the future demand for students. Furthermore, the industry's need for data analysis skills serves as the foundation for talent development in higher education[11]. Universities should be encouraged to foster the integration of industry and education, enhance the depth of collaboration between academia and businesses, and facilitate the development of data analysis skills among students.

Acknowledgment

The authors gratefully acknowledge the support of the Philosophy and Social Science Planning Project of the Guangdong Province of China (GD22XJY16) and the support of the 2020 Guangxi Vocational Education Reform Research Project (GXGZJG2020A006).

References

1. Melisa, Muñoz-Murillo, Pilar, B. Álvarez-Franco, Diego A. (2020) Restrepo-Tobón. The role of cognitive abilities on financial literacy: New experimental evidence, Journal of Behavioral and Experimental Economics, 84: 101482.
2. Persaud, A. (2021). Key competencies for big data analytics professions: A multimethod study. Information Technology & People, 34(1), 178-203.
3. Zhu, Y., Song, Z., Li, S., Wu, X. (2023). Research on the coupling measurement of medical higher education and health resource allocation in the context of healthy China.
4. Geng, Y., Huang, F. (2022). Coupling coordination between higher education and environmental governance: Evidence of western China. Plos one, 17(8), e0271994.
5. Huang, B. (2021, June). Research on the Promotion and Implementation of the "1+ X" Certificate System. In 2021 International Wireless Communications and Mobile Computing (IWCMC) (pp. 1608-1611). IEEE..
6. Kobets, V., Koval, S. (2023). Gap-filling model between the labor market requirements and the standards of educational programs for business analysts in Ukraine. Applied Aspects of Information Technology, 2(6): 139-150.
7. Ng, P. M., Chan, J. K., Wut, T. M., Lo, M. F., Szeto, I. (2021). What makes better career opportunities for young graduates? Examining acquired employability skills in higher education institutions. Education+ Training, 63(6): 852-871.
8. Krücken, G. (2021). Multiple competitions in higher education: a conceptual approach. Innovation, 23(2), 163-181.
9. Poláková, M., Suleimanová, J. H., Madzík, P., Copuš, L., Molnárová, I., Polednová, J. (2023). Soft skills and their importance in the labor market under the conditions of Industry 5.0. Heliyon, 9(8).
10. Riyani, E. R., Rizal, M. S., Ananda, R., Mufarizuddin, M., Aprinawati, I. (2022). Make a Match Type of Cooperative Learning Model to Social Skills of Elementary School Students. Edumaspul: Jurnal Pendidikan, 6(2): 1435-1439.
11. McCartney, S., Murphy, C., Mccarthy, J. (2021). 21st century HR: a competency model for the emerging role of HR Analysts. Personnel Review, 50(6): 1495-1513.