

# Research on the Current Application Status of Big Data Technology in Higher Vocational Education

Yongling Liu\*, Xuejun You, Taizhi Lv

School of Information Engineering, Jiangsu Maritime Institute, Nanjing 211170, China.

\*Corresponding Author: 594697472@qq.com

## Abstract

**Big data technology has extensive prospects in the field of higher vocational education, as it can assist institutions in gaining a better understanding of students and faculty, enhancing educational quality, and optimizing campus management. Researchers and educational institutions are actively exploring this area to address challenges and opportunities in the education sector. This paper aims to investigate the current application status of big data technology in higher vocational education and its potential impacts. Through a literature review, it examines the application of big data technology in areas such as student learning behavior analysis, the construction of talent development systems, and faculty scientific research in higher vocational education.**

## Keywords

**Education big data , Higher vocational education, Big data technology,Pandas, Data visualization.**

## 1. Introduction

With the advent of the information age, higher vocational education is facing unprecedented challenges and opportunities. The increase in the number of students, the abundance of educational resources, and the rise of online education have made teaching and management in higher vocational institutions more complex. At the same time, traditional methods of educational management have revealed shortcomings, and there is an urgent need for new approaches to improve the quality of education and management efficiency. The emergence of big data technology provides higher vocational institutions with new tools and opportunities. By collecting, storing, and analyzing vast amounts of data, colleges can better understand students' needs, optimize course design, enhance faculty performance, and engage in intelligent campus management [1].

The National Vocational Education Reform Implementation Plan (Vocational Education 20 Measures) proposes that vocational education should adapt to the development needs of Internet + vocational education and use modern information technology to improve teaching methods. The deep integration of new-generation information technologies, especially big data, the Internet of Things, and artificial intelligence, with education makes it possible to deeply track and quantify the learning process, collect and aggregate various data in educational scenarios, and even other cross-disciplinary data. The development and reform of the education sector in China is facing unprecedented opportunities and challenges, and the integration and application of big data in education and teaching are inevitable requirements of the times. As the point of convergence for new-generation information technology integration applications, applying big data technology to information technology-related majors in higher vocational education is a practical requirement and a future trend in the development of vocational talent. It has practical significance and can effectively promote the development of vocational education towards personalization, scientificity, and intelligence. As

the next natural resource of humanity, big data is becoming a powerful driver of organizational innovation, industrial upgrading, and economic development [2]. It has strong application demand and broad prospects in many industries and sectors of modern society. The education sector is an important application field for big data. As of June 15, 2023, there are a total of 3,072 higher education institutions in the country, with 2,820 regular higher education institutions, including 1,275 undergraduate institutions and 1,545 higher vocational (technical) colleges. There are also 252 adult higher education institutions. The number of college students has exceeded 44.3 million, accounting for more than 20% of the total global higher education scale. The enormous scale generates a wealth of educational basic data and associated data, making it one of the core data for the country, and it should be fully utilized in the process of advancing modern education.

In order to gain a deeper understanding of the application status of education big data in higher vocational education, this paper summarizes the main achievements and development trends in the field of education big data in the higher vocational domain, including recent research directions and key findings. This provides support for strengthening the deep integration of big data technology and concepts with the cultivation of vocational talents.

## 2. Educational Big Data Literature Analysis Based on CNKI

### 2.1. Analysis Methodology

To achieve automated analysis of literature data, this paper constructs a framework based on data crawling technology to collect, analyze, store, and visualize data from CNKI (China National Knowledge Infrastructure). The framework consists of data collection layer, data computation layer, data storage layer, and data visualization layer. The data source for this study is literature in CNKI with theme related to education big data and higher vocational education. The collected information includes paper titles, authors, journal names, publication dates, citation counts, download counts, abstracts, keywords, funding sources, classification codes, page numbers, and indexing information. Data scraping is accomplished using the Request and Beautiful Soup components. To ensure data quality and analysis accuracy, it is necessary to clean the literature data obtained from CNKI. Data cleaning primarily involves deduplication, handling missing values, standardizing formats, and handling outliers. Statistical and analytical work on literature feature data helps scholars keep up with academic progress, understand research hotspots and trends, identify research gaps and blanks, promote academic exchange, improve research efficiency, and support relevant decision-making. Pandas is an open-source data analysis library that provides high-performance, user-friendly data structures and data manipulation tools for data cleaning and analysis [3]. In this paper, literature data is statistically and analytically processed using the Pandas library, and the results are stored in a MySQL database. Data visualization is implemented using a front-end and back-end separation model. Front-end and back-end separation is a modern web application architecture in which the front-end and back-end are developed and deployed as two independent entities. The front-end is responsible for the user interface and experience, while the back-end handles data processing, business logic, and interaction with the database.

### 2.2. Analysis Results

The number of publications is one of the important indicators for assessing the level of attention to a research field, and the annual analysis of the number of publications can reflect changes in the level of attention to that research field. As shown in Figure 1, research on big data in higher vocational education began in 2014, and its research trend has been steadily increasing year by year. By 2029, the number of related research literature has increased to 183 papers, demonstrating the popularity of research in the field of big data in higher

vocational education, attracting a large number of scholars to engage in related research. After 2019, the related research gradually stabilized, with an annual quantity of around 100 papers.

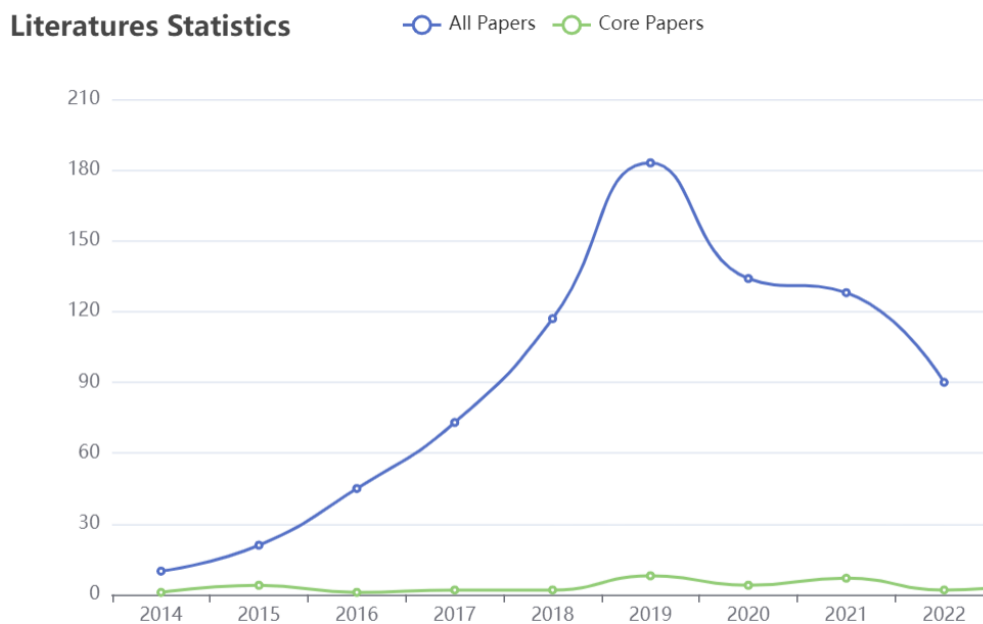


Figure 1: Annual number of literatures issued

The statistical analysis of publication journals helps researchers gain a better understanding of publication trends within the academic field and select appropriate journals for paper submissions. Journals in this field exhibit a considerable degree of dispersion, but there are several journals with relatively higher numbers of publications. For example, the Modern Vocational Education journal has published 30 papers. As shown in Figure 2, by examining journals with more than 10 publications, authors can gain a clearer understanding of which journals have more influence or are more popular in the field of big data research in higher vocational education.

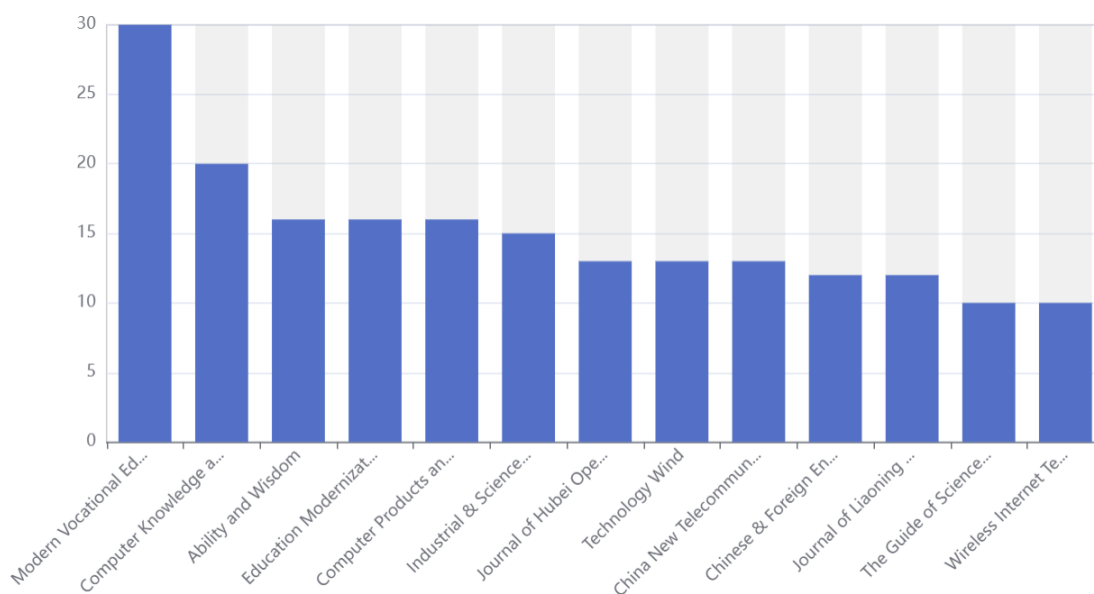


Figure 2: Literature distribution from different journals

Statistical analysis of author affiliations is beneficial for understanding the authors' identities and the distribution of research capabilities. Based on the statistics, the first authors of the 838 publications come from 427 different institutions, indicating a relatively dispersed distribution of research efforts. When looking at the publication volume of each institution, there are 15 institutions with a publication volume of 5 or more papers, indicating their strong potential and capabilities in the field of educational big data. The number of publications for each researcher (based on first authorship) is relevant to assessing the core authors in the vocational education research field. Often, a specific research field is driven by a small number of core authors who contribute to academic innovation and disciplinary development. The highest number of publications by a single scholar is 3 papers, with only a small number of authors publishing 2 papers. According to the Price Core Author Calculation Formula [4] popular in international literature, a core group of authors in the field of big data research in higher vocational education has not yet formed.

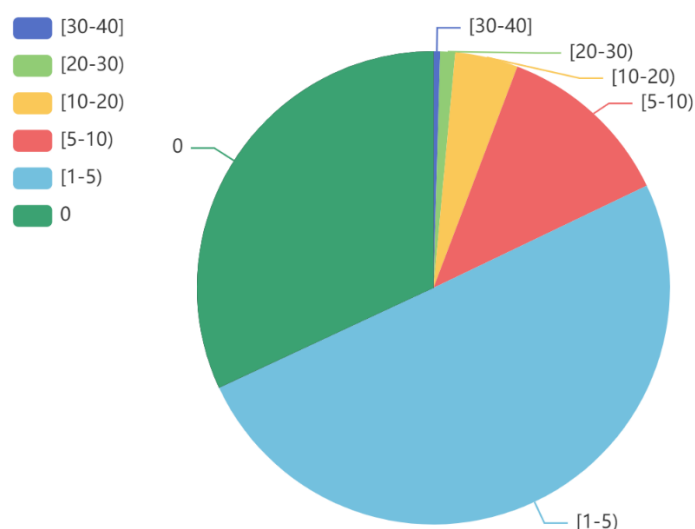


Figure 3: Citation frequency statistics of literatures

### 3. Scenario Analysis of Big Data Technology in Higher Vocational Education Teaching

#### 3.1. Research on Job Analysis and Recommendations Based on Big Data Technology

The biggest challenge facing talent development in higher vocational colleges is the inconsistency with the structure of societal talent demand [5]. From the supply side, the number of graduates from higher vocational colleges is increasing every year, leading to rising employment pressure. Many graduates from these institutions are encountering difficulties in finding employment. However, from the demand side, many enterprises are facing challenges in recruiting suitable technical personnel. This mismatch between supply and demand reflects an imbalance in the correspondence of majors and the supply-demand ratio in vocational education. In order to address the issue of the misalignment between talent development and societal demand, higher vocational colleges should closely follow industry developments, actively understand the employment requirements of technical talents, and nurture vocational skills. Job recruitment websites provide an effective and direct means of understanding the demands for technical talents [6]. Job postings on recruitment websites typically include skill

and qualification requirements for candidates. Establishing an automated job analysis platform to collect, store, analyze, and visualize job information has significant practical significance.

Using big data technology to mine a vast amount of job recruitment information has become an effective way to understand the industry's demand for talents. With the gradual maturation of web text mining technology, an increasing number of researchers and educators are applying text mining techniques to the field of education. Marjit etc. employs distributed web crawling technology to retrieve educational resources from the internet. After undergoing cleaning, noise reduction, and refinement processes, these resources are saved and utilized for educational purposes [7]. Liu etc. uses TF-IDF, word2vec, and K-means algorithms to analyze the requirements of positions related to big data. The aim is to help companies identify high-quality candidates [8]. Shen etc. use big data technology to analyze recruitment information on 51job, and the analysis results contribute to professional development [9]. Schedlbauer etc. crawl online recruitment data, and then, through data cleaning and visualization processes, uncover patterns within the data and provide job matching recommendations [10].

### **3.2. Research on Teaching Behavior Analysis Based on Big Data Technology**

Analyzing the data related to student learning stored in the academic system and using it to enhance teaching is essential. The demands of the job market are constantly evolving, and to better prepare students for these changing demands, schools may adjust their course offerings in real-time based on industry needs. Faced with new courses, students with lower learning abilities may struggle to keep pace with the teacher. Therefore, analyzing academic data in real-time to understand students' learning progress and making real-time adjustments to teaching schedules becomes particularly important [11].

Many scholars have conducted extensive analysis of educational data to improve the quality of talent development. Wang Yan et al., based on the evaluation data of all courses within a university semester, established a machine learning TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) model using five indicators to extract teaching characteristics. They demonstrated statistically significant correlations in course performance [12]. Lin etc. extracted raw data from a higher vocational college's academic affairs system and conducted mining and analysis using Python. This analysis serves as the basis for decisions related to the configuration of majors and courses, as well as the development of teaching strategies, ultimately enhancing the educational and managerial standards of the higher education institution [13]. Hajra etc. conduct a study in which they extracted a set of features from clickstream data in virtual learning environments. They utilized a deep artificial neural network to predict potentially at-risk students, offering measures for early intervention [14].

### **3.3. Research on the Analysis of Teacher Research Behavior Based on Big Data Technology**

Academic papers are important indicators for assessing the research capabilities of higher education institutions. They represent the academic viewpoints and scientific research achievements of teachers, serving as crucial indicators reflecting the research capabilities of vocational colleges and universities. For first-class universities and high-level undergraduate institutions, the focus is primarily on the quantity of papers indexed in SCI (Scientific Citation Index) or EI (Engineering Index). For vocational colleges and applied undergraduate programs, most of the teachers' papers are published in Chinese journals, which are typically indexed by the China National Knowledge Infrastructure (CNKI). Therefore, evaluating the research capabilities of a vocational college based on papers indexed in CNKI is feasible [15]. Yu, based on the 2021 domestic paper, fund project, and patent data of 14 selected transportation-related higher vocational colleges in the first round of the 'Double High Plan,' conducted numerical statistics, bibliometric analysis, and comparative analysis. The analysis of literature data is

conducive to the introspection, reflection, transcendence, and innovation of transportation-related national 'Double High Plan' higher vocational colleges. It sets an example for other transportation-related higher vocational colleges to write a new chapter in the research journey of higher vocational education in the new era, better serving the national strategy and regional economic and social development [16]. Ma etc. conduct an analysis of the research performance of 72 national demonstrative higher vocational colleges in China using the DEA (Data Envelopment Analysis) method. The analysis results indicate that the research performance of these demonstrative higher vocational colleges in China is non-DEA efficient, with significant differences among institutions. The non-DEA efficient colleges mainly face issues related to input redundancy [17]. Liu etc. conduct a bibliometric and visualization analysis based on academic papers, research projects, patents, and other research achievements data from 10 'Double High A-Level' vocational colleges in 2019, obtained from the CNKI database. The analysis highlights the exemplary role played by these 'Double High A-Level' colleges in vocational education research, leading other vocational colleges to serve national strategies, promote deep integration with industry-leading enterprises, gradually form a community of shared destiny between colleges and enterprises, and write a new chapter in the research journey of high vocational education with Chinese characteristics in the new era [18].

#### 4. Conclusion

The construction of educational big data has risen to a national strategic level. Since the 18th National Congress of the Communist Party of China, the central government and the State Council have attached unprecedented importance to information technology. Policies such as the Internet Plus Action Plan, the Action Plan for Promoting Big Data Development, and the Development Plan for the New Generation of Artificial Intelligence have been intensively issued. Among them, the Action Plan for Promoting Big Data Development explicitly mentions the construction of education and culture big data, making educational big data a national strategy and ushering in a major historical development opportunity for educational big data construction. Currently, as the informationization level of various universities continues to improve, a shared data platform has gradually been formed, including data such as campus smart cards, student grades, educational attendance, library borrowings, access control, and network access. This platform provides a data foundation for analyzing, processing, and mining potential data patterns in the campus big data environment, researching student behavior characteristics, analyzing student behavioral patterns, and enhancing school management and decision-making capabilities.

In the future, there are many research directions and trends for the application of big data in higher vocational colleges:

- Deep Learning and Artificial Intelligence: With the development of deep learning and artificial intelligence technologies, higher vocational colleges can use these technologies to better analyze and mine educational data, providing more intelligent educational services.
- Data Security and Privacy Protection: As the scope of big data application expands, data security and privacy protection become important issues. Future research needs to focus on how to protect the personal information of students and teachers and prevent data breaches.
- Interdisciplinary Research: Big data applications require interdisciplinary research, and future research trends may include cooperation across multiple fields such as education, computer science, and statistics.

- Personalized Education: Big data can enable personalized education by analyzing individual student data and tailoring educational experiences to their specific needs and learning styles.

Overall, the application of big data in higher vocational education holds great potential for improving the quality of education and optimizing institutional management. It will continue to play a crucial role in shaping the future of higher vocational education.

## Acknowledgements

This work was financially supported by the funding the Philosophy and Social Science Research Project of the Jiangsu Higher Education Institutions of China (2020SJA0730), the 13th Five-Year Plan Project of Education Science in Jiangsu Province(B-a/2020/03/26), the Fundamental Computer Education Research Project (2021-AFCEC-285) of the Association of Fundamental Computer Education Research Project of China, and the Excellent Teaching Team for QingLan Project of the Jiangsu Higher Education Institutions of China (Big Data Technology Teaching Team with Shipping Characteristic).

## References

- [1] Poloju, Satya Nagendra Prasad. "Applications Of Big Data Technology And Cloud Computing In Smart Campus." *International Journal of Creative Research Thoughts (IJCRT)*, ISSN: 2320 2882: 840-844.
- [2] Dubey, Rameshwar, et al. "Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture." *British Journal of Management* 30.2 (2019): 341-361.
- [3] Stančin, Igor, and Alan Jović. "An overview and comparison of free Python libraries for data mining and big data analysis." 2019 42nd International convention on information and communication technology, electronics and microelectronics (MIPRO). IEEE, 2019: 977-982.
- [4] Price, Derek de Solla. "A general theory of bibliometric and other cumulative advantage processes." *Journal of the American society for Information science* 27.5 (1976): 292-306.
- [5] Lv, Taizhi, Juan Zhang, and Yingying Chen. "Job Analysis System Based on Spark Platform." in proceedings of of the 2020 3rd International Conference on Financial Management, Education and Social Science, 2020: 307-311.
- [6] Tiru, Laurentiu, and Mădălina Mohorâta. "Attitudes toward e-recruitment. An explorative study in Romania." *Academicus International Scientific Journal* 11.21 (2020): 73-88.
- [7] Marjit, Ujjal, and Prabhakar Kumar. "Towards a decentralized and distributed framework for open educational resources based on ipfs and blockchain." in proceedings of of the 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE, 2020: 1-6.
- [8] R. L. Liu, W. H. Ye, R. Q. Gao, et al. "Text Clustering Research Based on big data job demand. " *Data Analysis and Knowledge Discovery*, 2017,1 (12): 32-40
- [9] Shen, Jingjing, and Shuyan Yu. "Visual Analysis of Big Data Related Job Recruitment Information Based on 51job." 3rd International Conference on Internet, Education and Information Technology (IEIT 2023). Atlantis Press, 2023: 211-225.
- [10] Schedlbauer, Jürgen, Georgios Raptis, and Bernd Ludwig. "Medical informatics labor market analysis using web crawling, web scraping, and text mining." *International Journal of Medical Informatics* 150 (2021): 104453.
- [11] Fischer, Christian, et al. "Mining big data in education: Affordances and challenges." *Review of Research in Education* 44.1 (2020): 130-160.
- [12] [Wang Yan, Zhou Kai, and Shen Shoufeng. "Mining and Clustering Analysis of Academic Affairs Big Data Based on Entropy Weighting Method." *Journal of Zhejiang University of Technology* 51.1 (2023): 84-87.

- [13] Lin Dezhi, Zhang Liu. "Research on BI Portrait of Higher Vocational College Academic Affairs System Based on Python." *Electronic Components and Information Technology*, 2022, 6(01): 180-181.
- [14] Waheed, Hajra, et al. "Predicting academic performance of students from VLE big data using deep learning models." *Computers in Human behavior* 104 (2020): 106189.
- [15] Lv Taizhi, Zhang Juan, and Rui Fengfeng. "An Evaluation System for Research Capabilities of Higher Vocational Colleges Based on CNKI Data. " *Journal of Beibu Gulf University* 035.011 (2020): 71-75, 84.
- [16] Yu Youhui. "Statistical Analysis of Research Achievements in Transportation-Related National 'Double High Plan' Higher Vocational Colleges: Based on 2021 Data from CNKI." *Industrial Technology and Vocational Education*, 2023, 21(03): 121-124.
- [17] Ma Xinyue, Shi Weiping, Wang Wan. "Research on Research Performance of 72 National Demonstrative Higher Vocational Colleges: A Literature Analysis Based on DEA Model." *China Higher Education Research*, 2021(Z1): 48-52.
- [18] Liu Xia, Chen Lei, Liu Kang, et al. "Analysis of Research Achievements in Double High A-Level Colleges in 2019: Based on CNKI Data." *Journal of Higher Vocational Education (Journal of Tianjin Vocational University)*, 2021, 30(03): 51-57.