ISSN 2723-6471

Data Envelopment Analysis of Scientific Research Performance for Higher Vocational Colleges

Lin Zhou[1], Sutthiporn Boonsong[2, *], Issara Siramaneerat[3, *], Thosporn Sangsawang[4], Pakornkiat Sawetmethikul[5]

[1] *Vocational Education Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110 Thailand*

[2] *Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110 Thailand*

[3] *Faculty of Liberal Arts, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110 Thailand*

[4] *Educational Technology and Communications Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand*

[5] *Department of Industrial Education Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand*

(Received: October 28, 2023; Revised: November 18, 2023; Accepted: December 23, 2023; Available online: January 29, 2024)

Abstract

This research aims to evaluate the scientific research performance of higher vocational colleges in Sichuan within the evolving landscape of data science. The study pursues two primary objectives: firstly, to assess the scientific research performance of these institutions using advanced methodologies such as Data Envelopment Analysis (DEA) and the Malmquist index models; secondly, to explore the intricate relationship between scientific research inputs and efficiency through the lens of Rough Set theory. The dataset comprises scientific research inputs and outputs from 30 higher vocational colleges, spanning the years 2019 to 2021. The findings underscore an overall positive trend in scientific research performance across the higher vocational colleges under examination. However, a nuanced analysis using DEA and Malmquist index models identified that only five institutions demonstrated robust performance during the specified period. Furthermore, the study delves into the influential factors affecting scientific research efficiency, revealing that internal expenditure on scientific research funds and the presence of senior and above professional teachers play pivotal roles. These insights are gleaned through the application of Rough Set theory, providing a unique perspective within the realm of data science. In conclusion, the research recommends strategic interventions to improve research management and resource allocation, emphasizing their role in enhancing efficiency and mitigating disparities among higher vocational colleges in Sichuan, particularly in the context of data science. The study adopts a holistic approach, employing an integrated model that combines DEA, Malmquist, and Rough Set theory for a comprehensive evaluation of research performance within the evolving landscape of data science.

Keywords: Scientific Research Performance, Data Envelopment Analysis, Malmquist Index Model, Rough Set Theory, Efficiency

1. Introduction

In January 2019, China's State Council issued the National Vocational Education Reform Implementation Plan, underscoring the crucial interplay between vocational education and general education in the broader context of educational reform and socio-economic development. Within this framework, higher vocational colleges were identified as key players in the regional innovation system [1]. Shifting the focus to the realm of data science, the Ministry of Education of China, in August 2022, released "Several Opinions on Strengthening Organized Scientific Research in Colleges and Universities to Promote High-level Self-reliance and Self-improvement," emphasizing the pivotal role of universities, now including data science applications, in enhancing China's scientific and technological prowess.

The substantial increase in state investment in research at colleges and universities has ignited discussions around evaluating research performance and optimizing the allocation of scientific and technological resources. In the context of data science, concerns arise about harnessing these resources effectively for advancing the field [2]. The research capabilities of higher vocational colleges, reflective of their contributions to national strategy and regional economic and social development, have gained significance in the context of data-driven insights and analytics [3]. Consequently,

***Corresponding author: Sutthiporn Boonsong, Issara Siramaneerat (sutthiporn_b@rmutt.ac.th, issara_s@rmutt.ac.th)**
DOI: https://doi.org/10.47738/jads.v5i1.166

 ISSN 2723-6471

research has become increasingly pivotal for the development of higher vocational colleges in the realm of data science [4].

Despite the growing importance of data science in research, higher vocational colleges in China still grapple with challenges such as unclear research direction, inadequate management, and insufficient investment in human and material resources specific to data science [5]. There is an observable inclination towards prioritizing quantity over quality in data science research achievements, leading to lower overall standards [6].

Motivated by these factors and acknowledging the unique challenges faced in the field of data science, this study aims to evaluate the scientific research performance of higher vocational colleges in Sichuan using DEA and Malmquist index models. Additionally, it seeks to delve into the influence of scientific research inputs, especially those related to data science, on research efficiency through the application of Rough Set theory.

2. Literature Review

Numerous scholarly investigations have consistently emphasized the versatility and effectiveness of the Data Envelopment Analysis (DEA) and Malmquist index models, particularly within the ever-evolving and dynamic landscape of data science. These models have been extensively examined and employed in various contexts for assessing resource utilization efficiency, but their specific applications within the nuanced domain of data science have become increasingly pivotal [7]–[12].

Within the expansive field of data science, research has elucidated the potential of the DEA model when applied to establish research input indicators encompassing a spectrum of resources, including human capital, financial investments, and material assets. Simultaneously, the incorporation of output indicators, measuring the quantity, quality, and economic benefits derived from scientific research achievements, has proven instrumental in explaining the multifaceted changes in scientific research performance, especially in the context of data science [8], [13]–[16]. Despite this, there is a noticeable void in the current scholarly literature concerning the application of diverse models, specifically tailored to address the intricacies of data science, for the comprehensive evaluation of scientific research performance.

Moreover, in the interconnected realms of higher education and economic growth, a relationship that extends its influence to the data science landscape, empirical studies have increasingly acknowledged the symbiotic nature of these two entities [17]. In the context of Sichuan Province, where higher vocational education development is intricately interwoven with the region's economic and social progress, there is a growing need to align scientific research efforts with the demands and advancements in data science [18]. With Sichuan Province proudly boasting 82 vocational colleges and ranking fifth among the 31 provinces in China, it stands as a noteworthy focal point for exploring the intersection of data science and vocational education. Nevertheless, empirical evidence pertaining to the scientific research performance of higher vocational colleges in Sichuan Province, particularly within the intricate realm of data science, remains conspicuously scarce.

This research endeavors to bridge this notable gap in the literature, aspiring to provide a thorough examination and valuable insights into the scientific research performance of higher vocational colleges in Sichuan Province, with a specific emphasis on data science-related indicators. By adopting a holistic approach, encompassing DEA and Malmquist index models, along with the innovative application of rough set theory, this study aims to contribute substantively to the enhancement and optimization of scientific research performance in higher vocational institutions, catering specifically to the demands and intricacies of the burgeoning field of data science.

3. Method

3.1. Population and Sample

This study focuses on the scientific research landscape within the domain of data science, encompassing a comprehensive dataset derived from 82 higher vocational colleges in Sichuan Province. The sample, drawn from this broader population, specifically includes published data from a subset of 30 higher vocational colleges. These colleges are selected based on their availability of scientific research inputs and outputs, spanning the pivotal years from 2019

to 2021. It's worth noting that the sample size of 30 colleges in this study, though larger than those examined by previous scholars [19], [20], ensures a representative cross-section of diverse higher vocational colleges in Sichuan Province. The chosen institutions include comprehensive, engineering, political and legal, agricultural and forestry, medical, and art colleges, providing a rich and varied dataset that mirrors the diversity of higher vocational education in the region. The distribution of these institutions across 18 of the 21 prefecture-level cities in Sichuan Province further adds geographical diversity to the sample.

3.2. DEA Model

The application of the Data Envelopment Analysis (DEA) method in the context of data science research offers several distinct advantages [21]. First and foremost, the DEA method obviates the need for a priori selection of weight coefficients, thereby minimizing the impact of subjective biases introduced by evaluators. This is particularly crucial in the data science domain, where objectivity and impartiality are paramount. Additionally, the DEA method does not mandate the standardization of input and output data units, a flexibility that aligns seamlessly with the diverse and dynamic nature of data science research. Owing to these characteristics, the DEA method has gained widespread acclaim and applicability in efficiency research scenarios that involve multiple inputs and outputs.

The chosen Data Envelopment Analysis model for this research leverages the Constant Returns to Scale (CRS) model and the Variable Returns to Scale (VRS) model, collectively known as the BCC model. Recognizing the variability in returns to scale inherent in data science research, the research employs the DEA-BCC model to scrutinize the research efficiency of higher vocational colleges within Sichuan Province. The mathematical expression of the model is articulated as follows:

$$\begin{cases} \min\theta \\ s.t.\sum_{j=1}^{n}\lambda_j x_j + s_i^- == \theta x_0, i = 1,2,\cdots,\alpha \\ \sum_{j=1}^{n}\lambda_j y_j - s_r^+ = y_0, r = 1,2,\cdots,\alpha \\ \sum_{j=1}^{n}\lambda_j = 1, \lambda_j \geq 0 \\ s_i^- \geq 0, s_r^+ \geq 0 \end{cases} \quad (1)$$

In Formula (1), the symbols serve the following purposes:

n represents the count of evaluation units; x_j signifies the input variable, while y_j signifies the output variable; s_i^- denotes the input slack variable indicating redundancy; s_r^+ represents the output slack variable indicating relaxation; λ_j corresponds to the parameter variable of the decision-making unit; θ denotes the efficiency evaluation value. When the equality $\theta = 1$, $s_i^- = 0$, $s_r^+ = 0$ is satisfied, it indicates the validity of the DEA assessment for the decision unit.

3.3. Malmquist Index Model

The Malmquist index, initially introduced by Sten Malmquist in 1953, is a method employed to analyze total factor productivity. It provides a convenient approach for evaluating production efficiency alterations across two distinct periods before and after the decision-making unit [22]. This method dissects the shift in productivity by accounting for technological advancement and efficiency alteration between the two temporal phases. As such, in this study, the Malmquist index was applied to compute the dynamic variations in the scientific research performance of higher vocational colleges within the province of Sichuan. The model is expressed as follows:

$$M_0(x^{t+1}, y^{t+1}, x^t, y^t) = \left[\frac{D^t(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} \times \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^{t+1}(x^t, y^t)}\right]^{\frac{1}{2}} \quad (2)$$

In Equation (2), x^t and y^t denote the input and output values for the *t-th* period, respectively, while D^trepresents the distance function for period *t*. The magnitude of D^t plays a pivotal role in assessing alterations in the total factor productivity.

When $M_0 > 1$, it indicates an enhancement in the total factor productivity of the research subject during the *t-th* period. In contrast, when $M_0 = 1$, it indicates that the total factor productivity of the research subject remains invariant throughout the *t-th* period. Conversely, when $M_0 < 1$, it suggests a reduction in the total factor productivity of the research subject over the specified time frame.

 ISSN 2723-6471

3.4. Rough Set Theory

Rough set theory, constitutes a mathematical approach employed to assess the significance of impact factors. This technique primarily concerns classifying data characterized by incompleteness, ambiguity, or inaccuracy based on empirical observations. Building upon the investigations conducted by Liang et al. [23] and Wan et al. [24], it is evident that rough set theory excels in the handling and analysis of diverse forms of incomplete information, encompassing imprecision, inconsistency, and incompleteness. It can unveil concealed insights and expose latent patterns within datasets, thereby showcasing robust objectivity and credibility in data analysis. Rough set theory has found widespread application in multi-attribute comprehensive evaluation challenges.

In this paper, the importance and contribution of input indicators to scientific research efficiency are studied by using the rough set method. The objective is to identify vulnerabilities within research input among higher vocational colleges in the province of Sichuan. The main contents of rough set theory are as follows: $S = (U\ A\ V\ F)$ referred to as a decision table, where $U = \{u_1, u_2, \cdots, u_n\}$ is a non-empty finite set called the universe. $A = \{a_1, a_2, \cdots, a_n\}$ is a non-empty finite set comprising all attributes and V_a represents the value domain of all attributes a.

$f: U \times A \to V$ is an information function such that for every object $a \in A, x \in U$ there exists $f(x, a) \in V_a$.

Let there be two equivalence sets, U/B and U/Q.

$U / B = \{[x] | x \in U, [x]$ is the equivalence class on set U based on the equivalence relation $B\}$

$U / Q = \{[y] | y \in U, [y]$ is the equivalence class on set U based on the equivalence relation $Q\}$

The information entropy of U/B,denoted by $H(B)$.

$$H(B) = -\sum_{i=1}^{n} P(X_i) \ln P(X_i) \tag{3}$$

The conditional entropy of knowledge Q relative to knowledge B is denoted as $H(B|Q)$:

$$H(B|Q) = -\sum_{i=1}^{n} P(X_i) \ln P(X_i) \frac{p(y_i|X_i)\log\,[p(y_i|X_i)]}{log\left(\frac{U}{Q}\right)} \tag{4}$$

In the information system $S = (U\ A\ V\ F)$, the importance of $a \in A$ is defined as:

$$S_A(a) = |H(A|Q) - H((A - a)|Q) \tag{5}$$

If $S_A(a) = 0$, then the attribute a is removed from the attribute set, and the remaining attributes are denoted as $C = \{C_1, C_2, \cdots, C_t\}$. Next, the importance $S_C(C_i)$ of each attribute in is recalculated and then normalized.

$$\omega = \frac{S_C(C_i)}{\sum_{i=1}^{t} S_C(C_i)} \tag{6}$$

3.5. Research Variables

Drawing on the research conducted by the scholars mentioned earlier, this study has formulated a system of indicators for evaluating the scientific research performance of higher vocational colleges in Sichuan. The system is outlined in Table 1.

Table 1. Scientific Research Performance Evaluation Indicators System of Sichuan Higher Vocational Colleges

Main Indicators	Indicators	Definition of Indicators
Research Input Indicators	Internal Expenditure of Scientific Research Funds	Financial resources allocated within an organization for the purpose of conducting scientific research activities.
	Full-time Teachers	The educators who are employed by a university or college and are actively involved in teaching activities on a full-time basis within the academic institution.
	Senior and Above Professional Titles Teachers	Teachers with senior titles such as professor and associate professor.
	Library Collections	The total amount of paper documents in university libraries.

 ISSN 2723-6471

Research Output Indicators	Research Papers	Academic articles published in academic journals or at academic conferences.
	Research Projects at the Municipal Level and Above	Projects issued by government departments at or above the municipal level or research institutions recognized as equivalent to government departments at or above the municipal level.
	Vertical and Horizontal Project Arrival Funds	The received funds for vertical projects (scientific research projects approved by government departments at all levels, academic organizations and units and supported by national financial allocation) and horizontal projects (scientific research projects established by enterprises or individuals).
	Granted Patents	The patents that have been formally approved by the relevant departments.

4. Result and Discussion

4.1. The evaluation of the scientific research performance of higher vocational colleges in Sichuan using DEA and Malmquist index models.

For the analysis of scientific research performance among higher vocational colleges in Sichuan Province from 2019 to 2021, this study employed the DEA model in an input-oriented approach. DEAP 2.1 software was utilized for the static analysis. The detailed results are provided in Tables 2 and 3.

Table 2. Research Efficiency Values of Higher Vocational Colleges in Sichuan Province from 2019 to 2021

Firm	2019				2020				2021			
	Crste	Vrste	Scale		Crste	Vrste	Scale		Crste	Vrste	Scale	
1	1.000	1.000	1.000	-	1.000	1.000	1.000	-	1.000	1.000	1.000	-
2	0.583	0.628	0.928	drs	0.961	0.966	0.995	drs	1.000	1.000	1.000	-
3	0.458	0.469	0.977	irs	0.487	0.491	0.992	irs	0.637	0.647	0.985	drs
4	0.302	0.355	0.853	drs	0.463	0.465	0.994	irs	0.421	0.444	0.948	drs
5	1.000	1.000	1.000	-	0.904	1.000	0.904	drs	0.879	0.971	0.906	drs
6	0.471	0.472	0.997	drs	1.000	1.000	1.000	-	1.000	1.000	1.000	-
7	0.839	0.887	0.946	irs	0.328	0.384	0.854	irs	0.853	1.000	0.853	irs
8	0.389	0.391	0.994	drs	0.350	0.386	0.907	drs	0.224	0.240	0.930	drs
9	0.536	0.692	0.774	drs	0.508	0.737	0.689	drs	0.473	0.517	0.914	drs
10	1.000	1.000	1.000	-	1.000	1.000	1.000	-	1.000	1.000	1.000	-
11	1.000	1.000	1.000	-	1.000	1.000	1.000	-	1.000	1.000	1.000	-
12	0.700	0.791	0.884	drs	0.711	0.875	0.812	drs	0.911	1.000	0.911	irs
13	0.068	0.068	0.989	irs	0.092	0.095	0.973	irs	0.266	0.294	0.902	drs
14	0.935	0.951	0.983	drs	0.839	0.856	0.980	irs	0.887	0.970	0.914	irs
15	0.546	0.567	0.963	drs	0.545	0.689	0.791	drs	0.594	0.682	0.870	irs
16	0.883	0.894	0.988	drs	1.000	1.000	1.000	-	1.000	1.000	1.000	-
17	1.000	1.000	1.000	-	1.000	1.000	1.000	-	1.000	1.000	1.000	-
18	0.234	0.280	0.836	irs	0.171	0.174	0.983	drs	0.604	0.713	0.847	irs
19	0.948	1.000	0.948	drs	0.913	1.000	0.913	drs	1.000	1.000	1.000	-
20	1.000	1.000	1.000	-	0.827	1.000	0.827	drs	0.663	0.674	0.983	irs
21	0.795	0.808	0.984	drs	0.728	0.752	0.968	drs	1.000	1.000	1.000	-
22	0.951	1.000	0.951	drs	0.837	0.955	0.876	drs	0.988	1.000	0.988	drs
23	0.862	0.901	0.957	irs	0.731	0.780	0.937	irs	0.823	0.848	0.970	irs
24	0.162	0.183	0.888	drs	0.056	0.062	0.890	drs	0.076	0.086	0.891	drs
25	0.802	0.836	0.960	irs	0.805	0.824	0.978	irs	0.620	0.780	0.796	irs
26	1.000	1.000	1.000	-	1.000	1.000	1.000	-	1.000	1.000	1.000	-
27	0.658	1.000	0.658	irs	0.917	1.000	0.917	irs	1.000	1.000	1.000	-
28	0.942	1.000	0.942	irs	0.599	0.785	0.764	irs	0.731	0.754	0.970	irs
29	0.550	0.562	0.979	irs	0.514	0.521	0.985	irs	1.000	1.000	1.000	-
30	0.243	0.259	0.938	irs	0.227	0.259	0.876	irs	0.213	1.000	0.213	irs

Table 3. Comprehensive Analysis of Scientific Research Performance in Higher Vocational Colleges within Sichuan Province from 2019 to 2021

 ISSN 2723-6471

Dea Efficiency	Efficiency Eigenvector	2019	2020	2021	Average
TE (Technical Efficiency)	Number of colleges with TE=1	7	7	12	8
	Average	0.695	0.684	0.762	0.714
PTE (Pure Technical Efficiency)	Number of colleges with PTE=1	11	11	16	12
	Average	0.733	0.735	0.821	0.765
SE (Scale Efficiency)	Number of colleges with SE=1	7	7	12	8
	Average	0.944	0.927	0.926	0.932
RE (Returns to Scale)	Increasing number of colleges	10	11	10	10
	Decreasing number of colleges	13	12	8	11
	Constant number of colleges	7	7	12	8

Based on the data presented in Tables 2 and 3, it is evident that the average technical efficiency of higher vocational colleges in Sichuan Province is relatively low, with 36% of these colleges experiencing decreasing returns to scale. This outcome indicates that the overall research and technical efficiency of higher vocational colleges in Sichuan Province during the period from 2019 to 2021 remains at a suboptimal level.

Technical efficiency, which assesses the effectiveness of research output in higher vocational education relative to input resources, examines how efficiently universities utilize available resources to maximize research outcomes. The average technical efficiency for higher vocational colleges in Sichuan Province in 2019, 2020, and 2021 was 0.695, 0.684, and 0.762, respectively. While these values show an overall improvement trend, they still fall within the lower range, indicating the need for further improvement in resource allocation and utilization efficiency among higher vocational colleges in Sichuan Province. Only 5 out of the total higher vocational colleges achieved continuous DEA effectiveness over three years, representing 16% of the total. These colleges have successfully optimized their resource allocation, resulting in the best input-output performance in scientific research. Despite the progressive improvement in technical efficiency for most colleges, significant room for advancement still exists.

Pure technical efficiency, which evaluates a university's output relative to input resources while accounting for scale effects, examines a university's ability to leverage technology and innovation within existing resource utilization to maximize research output. The mean pure technical efficiency for higher vocational colleges in Sichuan Province was 0.733, 0.735, and 0.821 for 2019, 2020, and 2021, respectively. Although these values display an upward trend, they remain relatively moderate, indicating ample room for improvement in research management, resource utilization, and technological innovation within higher vocational colleges in Sichuan Province. Six colleges consistently achieved effectiveness in pure technical efficiency over three years, highlighting their relative advancement in technological innovation and proficient management.

Scale efficiency assesses whether a university operates at an appropriate scale to achieve maximum research output. It compares a university's actual scale to the theoretically optimal scale to determine whether it effectively leverages its scale advantage. The average scale efficiency for higher vocational colleges in Sichuan Province was 0.944, 0.927, and 0.926 for 2019, 2020, and 2021, respectively, all exceeding 0.9. This suggests that the research input scale of higher vocational colleges in Sichuan Province is reasonable, and they have achieved commendable efficiency levels. The average scale efficiency surpasses the average pure technical efficiency, indicating that research management and technical capabilities are the primary factors constraining the scientific research performance of higher vocational colleges. These results indicate that some colleges can enhance their scientific research performance by expanding their input scale.

Regarding the scale return index, higher vocational colleges in Sichuan Province exhibit varying trends. Some colleges experienced increasing scale returns, while others faced decreasing scale returns. These findings highlight the importance of adjusting the allocation of research resources based on individual circumstances. Colleges with increasing scale returns may benefit from increasing research investment, while those with decreasing scale returns can optimize overall efficiency by reducing resource inputs.

 ISSN 2723-6471

This study further employs the Malmquist index model and utilizes DEAP 2.1 software to conduct a dynamic analysis of research performance among higher vocational colleges in Sichuan Province from 2019 to 2021, with the results presented in Tables 4 and 5.

Table 4. Malmquist Index and Its Components of Scientific Research Performance Evaluation of Higher Vocational Colleges in Sichuan Province from 2019 to 2021

Firm	Effch (Efficiency Change)	Techch (Technical Change)	Pech (Pure Technical Efficiency) Change)	Sech (Scale Change)	Tfpch (Total Factor Productivity Change)
1	1.000	0.833	1.000	1.000	0.833
2	1.310	0.902	1.262	1.038	1.182
3	1.179	1.095	1.175	1.004	1.291
4	1.180	1.001	1.119	1.054	1.181
5	0.938	1.037	0.985	0.952	0.973
6	1.458	1.104	1.455	1.002	1.609
7	1.009	0.854	1.062	0.950	0.861
8	0.758	1.123	0.784	0.967	0.851
9	0.939	0.983	0.864	1.086	0.923
10	1.000	1.010	1.000	1.000	1.010
11	1.000	0.701	1.000	1.000	0.701
12	1.141	0.910	1.124	1.015	1.038
13	1.983	1.161	2.077	0.955	2.302
14	0.974	0.893	1.010	0.964	0.869
15	1.043	0.952	1.097	0.951	0.992
16	1.064	1.125	1.058	1.006	1.197
17	1.000	0.898	1.000	1.000	0.898
18	1.605	1.023	1.595	1.006	1.641
19	1.027	1.144	1.000	1.027	1.175
20	0.814	1.073	0.821	0.992	0.874
21	1.121	0.910	1.112	1.008	1.020
22	1.019	1.074	1.000	1.019	1.095
23	0.977	1.177	0.970	1.007	1.149
24	0.685	0.952	0.684	1.002	0.652
25	0.879	1.077	0.966	0.910	0.948
26	1.000	1.162	1.000	1.000	1.162
27	1.233	1.327	1.000	1.233	1.636
28	0.881	1.223	0.868	1.015	1.078
29	1.348	1.381	1.333	1.011	1.862
30	0.935	0.999	1.963	0.476	0.934
Average	**1.058**	**1.027**	**1.080**	**0.980**	**1.086**

Table 5. Malmquist Index and Its Decomposition of Scientific Research Performance of Higher Vocational Colleges in Sichuan Province from 2019 to 2021 (Each Year)

Year	Effch	Techch	Pech	Sech	Tfpch
2019-2020	0.955	1.198	0.974	0.981	1.145
2020-2021	1.171	0.879	1.197	0.979	1.030
Average	1.058	1.027	1.080	0.980	1.086

The analysis of Tables 4 and 5 reveals that from 2019 to 2021, the average total factor productivity (TFP) of 30 higher vocational colleges in Sichuan Province was 1.086, indicating an overall TFP growth of 8.6%. Notably, the trend in pure technical efficiency alteration aligns with the trend in technical efficiency alteration, which, in turn, corresponds with the trend in total factor productivity alteration. This alignment highlights the substantial influence of pure technical efficiency on the effective allocation and utilization of research resources within higher vocational colleges in Sichuan Province.

In terms of developmental dynamics, 18 colleges exhibited a total factor productivity (TFP) exceeding 1, with the most rapidly advancing institution experiencing a remarkable 130% increase in TFP. In contrast, 12 colleges had a TFP

ISSN 2723-6471

below 1, indicating that the overall research TFP of higher vocational colleges in Sichuan Province remains relatively modest, suggesting considerable room for improvement. During the same period, the average technical efficiency rose by 5.8%, while the average technological progress increased by 2.7%. This underscores the central role of research management and resource utilization efficiency within higher vocational colleges in Sichuan Province in elevating overall efficiency, with technological progress playing a secondary role.

Furthermore, the average pure technical efficiency increased by 8%, while the average scale efficiency declined by 2%. This observation suggests that the enhancement in technical efficiency is primarily driven by the increase in pure technical efficiency, emphasizing the significance of improving research management and resource utilization efficiency as key drivers of improved research efficiency.

In the annual analysis, it was found that from 2019 to 2020, the total factor productivity of higher vocational colleges in Sichuan Province surged by 14.5%. This period saw a decrease of 4.5% in the technical efficiency index and a simultaneous rise of 19.8% in the technological progress index. This underscores the pivotal role of advancements and innovations in research practices in augmenting the overall research efficiency of higher vocational colleges in Sichuan Province during this period. From 2020 to 2021, the total factor productivity of higher vocational colleges increased by 3%, with a concurrent increase of 17.1% in the technical efficiency index and a decline of 12.1% in the technological progress index. This highlights the primary role of research management practices and resource utilization efficiency in enhancing scientific research performance during this period. Consequently, higher vocational colleges in Sichuan Province have significant potential to elevate their scientific research performance by improving research technology management, enhancing resource utilization, and fortifying research innovation.

4.2. The result of studying the effect of scientific research inputs on scientific research efficiency using Rough Set theory

Based on the aforementioned analysis, it is evident that there is variability in the scientific research efficiency among higher vocational colleges in Sichuan. This stage employs Rough Set theory to investigate the impact of scientific research input indicators on scientific research performance, with the aim of revealing the relevance of various investment indicators to scientific research performance.

In accordance with Rough Set theory, the analyzed data in this stage includes technical efficiency, technical progress efficiency, and total factor productivity obtained from the Malmquist index model, as well as research input data from 30 higher education colleges for the years 2019 to 2021. The data is then discretized using a four-level classification method. In this method, the value is assigned 4 if it exceeds the mean plus 2 standard deviations, 3 if it is greater than the mean but less than the mean plus 2 standard deviations, 2 if it is less than or equal to the mean minus 2 standard deviations, and 1 if it is equal to or less than the mean minus 2 standard deviations. Subsequently, the information entropy, importance, and contribution of each scientific research input indicator based on different decision attributes are calculated and presented in Table 6.

Table 6. Information Entropy, Importance and Contribution of Scientific Research Input Indicators in Higher Vocational Colleges of Sichuan Province

	Research Input Indicators	Information Entropy	Importance	Contribution
Based on technical efficiency	Internal expenditure of scientific research	0.556	0.269	0.404
	Full-time teachers	0.382	0.096	0.144
	Senior and above professional titles teachers	0.520	0.233	0.349
	Library collections	0.356	0.069	0.104
Based on technical progress efficiency	Internal expenditure of scientific research	0.585	0.094	0.384
	Full-time teachers	0.558	0.067	0.273
	Senior and above professional titles teachers	0.537	0.046	0.188
	Library collections	0.529	0.038	0.155
Based on total factor productivity	Internal expenditure of scientific research	0.644	0.221	0.577
	Full-time teachers	0.428	0.0056	0.015
	Senior and above professional titles teachers	0.556	0.133	0.348
	Library collections	0.446	0.023	0.060

 ISSN 2723-6471

As presented in Table 6, we have obtained the information entropy, importance, and contribution values of four research input indicators concerning technical efficiency, technical progress efficiency, and total factor productivity. The results reveal that all four research input indicators exert an influence on the technical efficiency of higher vocational colleges. Specifically, "Internal expenditure of scientific research funds" and "Senior and above professional titles teachers" exhibit strong associations with technical efficiency, contributing 40.4% and 34.9%, respectively. The combined contribution of these two factors amounts to 75.3%, suggesting that differences in research technical efficiency among various higher vocational colleges are primarily determined by these two factors.

Furthermore, all four research inputs significantly impact the technical progress efficiency of higher vocational colleges, with the following contribution rankings: "Internal expenditure of scientific research funds," "Full-time teachers," "Senior and above professional titles teachers," and "Library collections." Interestingly, in contrast to the contribution ranking based on technical efficiency, "Full-time teachers" moved from the third to the second position, indicating a more prominent role in technical progress efficiency.

Lastly, the research findings demonstrate that all four research inputs contribute to the total factor productivity of higher vocational colleges. Among them, "Internal expenditure of scientific research funds" and "Senior and above professional titles teachers" have the highest contribution degrees, accounting for 57.7% and 34.8%, respectively. The combined contribution rate of these two factors is 92.5%. The lower contribution degrees of the other two inputs suggest that these two factors are primarily responsible for the differences in total factor productivity among various higher vocational colleges.

In summary, the research findings highlight the significant effects of internal expenditure for scientific research funds and senior and above professional teachers on scientific research efficiency. These two factors emerge as the main contributors to the variations in research input efficiency among different colleges.

5. Conclusion

This study undertakes a comprehensive analysis of the scientific research performance of 30 higher vocational colleges in Sichuan from 2019 to 2021, with a specific focus on the dynamic and static aspects within the realm of data science. By delving into this specific scope, the research aims to address the gap in existing literature concerning the overall scientific research performance of higher vocational colleges in Sichuan Province, thereby providing a nuanced understanding that can extend to other provinces in China seeking to study and enhance the scientific research performance of their higher vocational institutions.

The static analysis of the data uncovered that the average technical efficiency of higher vocational colleges in Sichuan province during the specified period was relatively low, with a consistent but modest increase over time. A notable finding was that only 5 out of the 30 higher vocational colleges achieved DEA efficiency over the three-year period, indicating that the majority, or 83%, of the colleges were still non-DEA efficient in the context of data science. Furthermore, 36% of the colleges experienced decreasing returns to scale, signaling sub-optimal overall scientific research performance and inefficient utilization of research resources in the realm of data science. The low mean values of pure technical efficiency also underscored a generally low level of internal research management capability and resource utilization efficiency within these vocational colleges, especially when considering data science-related research.

From a dynamic analysis perspective, the overall total factor productivity of higher vocational colleges in Sichuan province from 2019 to 2021 exceeded 1, indicating an overall improvement in comprehensive productivity. The overall scale efficiency was also greater than 0.9, suggesting that the overall research input scale in these colleges was relatively reasonable and achieved good efficiency, although there was room for improvement. However, the overall scale change index declined annually, indicating a decrease in the overall research input scale, somewhat hindering the improvement of research performance within the data science context. Additionally, the average technical efficiency was significantly higher than the average technical progress efficiency, highlighting the predominant role of research management and resource utilization efficiency in improving research performance, with research and technological innovation playing a secondary role, particularly in the field of data science.

These findings align with the results of other studies conducted in the western region of China, where the proportion of non-DEA efficient colleges reached a maximum of 64%, indicating lower overall scientific research performance than the national average [5]. Notably, Chengdu Aviation Vocational and Technical College, along with five other higher vocational colleges, achieved the benchmark level of research in the field of professional group construction, aligning with the list of higher vocational colleges identified in this study as having relatively high scientific research performance [20].

By employing the Rough Set method analysis within the realm of data science, this study identified "Internal expenditure of scientific research funds" and "Senior and above professional titles teachers" as the main factors contributing to differences in research input efficiency among different colleges. These findings align with the outcomes of previous studies, emphasizing the substantial influence of internal research funding allocation and the number of mid to senior-level faculty members with professional titles on university research performance, particularly in the context of data science [8], [25].

Based on these research findings, the following countermeasures and suggestions are proposed within the data science framework:

1) Optimize the strategy for allocating research resources: Tailor resource allocation strategies to colleges experiencing varying returns to scale, promoting sustained growth in research outputs. Emphasis should be placed on optimizing the allocation of resources for "Internal expenditure of scientific research funds" and "Senior and above professional titles teachers" to narrow the scientific research performance gap among colleges. Strengthening cooperation and communication between colleges, research institutions, and enterprises can facilitate resource sharing and complementarity, ultimately enhancing research performance, especially in the field of data science.
2) Continuously enhance the research management system of higher vocational colleges: Implement regular evaluations of scientific research performance and dynamically adjust research management systems to drive enhancements in data science-related scientific research performance standards.
3) Strengthen scientific and technological innovation: Higher vocational colleges in Sichuan province should strategically build high-quality research platforms based on their unique characteristics to serve regional economic and social development. Strengthening collaboration with enterprises, continuously improving research and technological progress, and efficiently transforming research outputs are crucial steps in this regard, particularly within the context of data science.

In conclusion, this study adopts an integrated approach, including the DEA-BCC model, Malmquist index model, and Rough Set theory, to comprehensively analyze the scientific research performance of higher vocational colleges, with a specific focus on the burgeoning field of data science. The research findings provide valuable insights into enhancing scientific research performance within higher vocational colleges, particularly within the data science domain, and mitigating inter-collegiate disparities in research performance. Consequently, the integrated models can be applied to scientific research performance analysis in universities and research institutions, offering specific innovation and generalization potential for future research endeavors in the field of data science and beyond.

6. Declarations

6.1. Author Contributions

Conceptualization: L.Z., S.B., I.S., and T.S.; Methodology: T.S.; Software: L.Z.; Validation: L.Z. and T.S.; Formal Analysis: L.Z. and T.S.; Investigation: P.S.; Resources: I.S.; Data Curation: P.S.; Writing Original Draft Preparation: P.S. and L.Z.; Writing Review and Editing: P.S. and L.Z.; Visualization: L.Z.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

 ISSN 2723-6471

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Wu, "Logical Requirements in the Construction of Scientific Research Platform in Higher Vocational Colleges from the Perspective of Science-education Integration," *Vocat. Tech. Educ.*, vol. 44, no. 10, pp. 25–30, 2023.

[2] J. Zhong and Z. Liu, "Research on the performance evaluation of scientific research in higher vocational colleges in China-Based on DEA analysis of school-level panel data from 2004 to 2015," *Contemp. Vocat. Educ.*, vol. 5, pp. 44–49, 2017, doi: 10.16851/j.cnki.51-1728/g4.2017.05.009.

[3] Y. Yu, "Statistical Analysis of Scientific Research Achievements in Higher Vocational Colleges Under the National ' Double High Plan' for Transportation —based on CNKI Data for 2021," *Ind. Technol. Vocat. Educ.*, vol. 3, pp. 121–124, 2023, doi: 10.16825/j.cnki.cn13-1400/tb.2023.03.028.

[4] H. Liu and H. Kuang, "Data Analysis of Scientific Research Achievements of National Higher Vocational Colleges in 2019 — — Based on CNKI Data," *Chinese Vocat. Tech. Educ.*, vol. 36, 2019.

[5] X. Ma, W. Shi, and W. Wang, "Research on the scientific research performance of 72 national demonstratives higher vocational colleges——literature Analysis Based on DEA Model," *China Univ. Sci. Technol.*, vol. Z1, pp. 48–52, 2021, doi: 10.16209/j.cnki.cust.2021.z1.010.

[6] W. Qiao, "Research on Performance Evaluation Model of Scientific Research in Higher Vocational Colleges," *High. Vocat. Educ. Explor.*, vol. 19, no. 2, pp. 24–29, 2020.

[7] J. Yan, "Spatiotemporal analysis for investment efficiency of China's rural water conservancy based on DEA model and Malmquist productivity index model," *Sustain. Comput. Informatics Syst.*, vol. 21, pp. 56–71, 2019.

[8] W. Cai and H. Yao, "The Input-output Performance and Influencing Factors of Scientific Research in"Double First - Class"Construction Universities: Analysis Based on DEA-Malmquist-Tobit Model," *Sci. Explor. Educ.*, vol. 40, no. 3, pp. 17–27, 2022.

[9] D. El Kadiri Boutchich, "Relevant Ingredients for Identifying Factors With Significant Impact on Research Structures Efficiency in Higher Education," *J. Educ.*, vol. 201, no. 3, pp. 248–255, 2021, doi: 10.1177/0022057420914910.

[10] C. Ibrahim and R. Fadhli, "Performance of Indonesia's World-Class University Efficiency with Bibliometrics (Scientific Strength) Approach and Data Envelopment Analysis," *Webology*, vol. 18, no. 1, pp. 32–50, 2021, doi: 10.14704/WEB/V18I1/WEB18003.

[11] X. Mo, Y. Chen, Y. Ren, and J. Zhu, "Research on research performance evaluation of universities in China based on DEA model," *Technol. Mark.*, vol. 30, no. 2, pp. 119–124, 2023.

[12] B. Soummakie and M. Wegener, "A Two-Stage Research Performance Assessment of Turkish Higher Education Institutions Using Data Envelopment Analysis and Beta Regression," *Academia.Edu*, 2021.

[13] L. A. Moncayo–Martínez, A. Ramírez–Nafarrate, and M. G. Hernández–Balderrama, "Evaluation of public HEI on teaching, research, and knowledge dissemination by Data Envelopment Analysis," *Socioecon. Plann. Sci.*, vol. 69, p. 100718, 2020, doi: 10.1016/j.seps.2019.06.003.

[14] L. P. Navas, F. Montes, S. Abolghasem, R. J. Salas, M. Toloo, and R. Zarama, "Colombian higher education institutions evaluation," *Socioecon. Plann. Sci.*, vol. 71, p. 100801, 2020, doi: 10.1016/j.seps.2020.100801.

[15] S. Liao, M. Ge, X. Su, and H. Lv, "Research on Research Efficiency Evaluation of Universities in Different Regions of China: An Empirical Analysis Based on the Categorized DEA Model," *China Univ. Sci. Technol.*, vol. Z1, pp. 38–42, 2021, doi: 10.16209/j.cnki.cust.2021.z1.008.

[16] Y. Wu, Y. Ba, and W. Zhang, "The Measurement and Influence Factors of Scientific Research Innovation Efficiency of

'First-class' University in China — An Empirical Study Based on Three-stage DEA Model," *Renmin Univ. China Educ. J.*, vol. 3, pp. 128–144, 2023.

[17] W. Shi, *Economics of Higher Education*. Science Press, 2004.

[18] Y. Lei, M. Xiong, J. Li, and K. Lou, "Study on the Relationship between Higher Vocational Education Input and Regional Economy Growth in Sichuan Province," *J. Huainan Vocat. Tech. Coll.*, vol. 2, no. 23, pp. 128–130, 2023.

[19] J. Zhang, "Analysis of the Current Situation and Management Strategies of Research Work in Vocational Colleges - A Case Study of Sichuan Vocational Colleges," *J. Taiyuan Urban Vocat. Coll.*, vol. 2, pp. 11–14, 2020, doi: 10.16227/j.cnki.tycs.2020.0075.

[20] Y. Xiao and S. Chen, "A comparative study of academic paper influence in Sichuan vocational colleges from the perspective of professional cluster development," *Neijiang Sci. Technol.*, vol. 43, no. 7, pp. 136–138, 2022.

[21] X. Xiao, Z. Zhong, Y. Wang, C. Zhang, and H. Wu, "Research on Energy Efficiency Evaluation of Urban Rail Transit Based on DEA-BCC Model," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2020, p. 12038. doi: 10.1088/1755-1315/435/1/012038.

[22] S. Malmquist, "Index numbers and indifference surfaces," *Trab. Estad.*, vol. 4, no. 2, pp. 209–242, 1953, doi: 10.1007/BF03006863.

[23] B. Liang and Z. Wang, "Research on agricultural land use data analysis and planning decisions based on rough set theory in Wuhan city," *Chinese J. Agric. Resour. Reg. Plan.*, vol. 39, no. 3, pp. 64–70, 2018.

[24] W. Wan, M. Wang, and X. Chen, "Research on multi-attribute measurement method based on rough set," *Comput. Mod.*, vol. 1, pp. 49–51, 2007.

[25] Z. Xu and G. Li, "Research on the Performance of High-Quality Scientific Research in Universities under Double First-Class Construction Based on Three-Stage DEA-TOBIT Model," *J. High. Educ. Res.*, vol. 46, no. 1, pp. 63–74, 2023.