Analysis of Teachers' Scientific Research Behavior Types Based on K-means Clustering: A Case Study of the School of Information Engineering Hainan Vocational University of Science and Technology

Darun Yang, Fan Yang
Hainan Vocational University of Science and Technology, Haikou, Hainan, China

Abstract: The pattern of teachers' scientific research behavior is of fundamental significance for university research management. Based on the data of scientific research projects at or above the municipal level undertaken by teachers from the School of Information Engineering, Hainan Vocational University of Science and Technology, between 2019 and 2025, this study employs the K-means clustering algorithm to classify the research behaviors of 40 teachers. By extracting seven dimensional features, including professional title rank, education background rank, total number of projects, and the quantities of four project categories, the data were standardized and clustered. The optimal number of clusters was determined to be K=5 using the Elbow Method and Silhouette Coefficient. Consequently, five types of teacher groups were identified: Teaching-Research Balanced Type, Novice Explorer Type, Local Service Type, Industry-Education Integration Type, and Research Leader Type. Significant differences were observed among these groups in terms of project quantity, project type preference, professional title, and education background structure, reflecting the diverse characteristics of teachers' research behaviors in vocational undergraduate universities. Based on the clustering results, differentiated support strategies are proposed, providing an empirical reference for research management in similar institutions.

Keywords: K-means clustering; Teachers' scientific research behavior; Vocational undergraduate; Research management

1. Introduction

Vocational undergraduate universities undertake the important mission of integrating science and education as well as integrating industry and education, with research capacity becoming a key indicator for measuring their educational quality [4]. As the main participants in scientific research activities, university teachers' research behavior patterns directly influence the institution's research output and disciplinary development [6]. However, current university research management often adopts uniform support strategies, neglecting the heterogeneity among teacher groups, which leads to inefficient resource allocation [1]. How to identify the typological characteristics of teachers' research behavior and subsequently implement precision management has become an urgent issue to address.

Data mining technology provides a new perspective for analyzing teachers' research behavior. K-means clustering, as a classic unsupervised learning method, can divide teachers into intrinsically homogeneous groups based on multi-dimensional features and has been widely applied in educational research [10]. Taking the School of Information Engineering at Hainan Vocational University of Science and Technology as an example, this study uses K-means clustering algorithm to classify teachers' research behavior based on detailed data of municipal-level and above research projects undertaken by teachers from 2019 to 2025, identifying the core characteristics of different teacher types to provide data support for optimizing research management.

2. Literature Review

2.1 Research on University Research Project Management

Using a Open University as a sample, Yan Bingshu conducted statistical analysis on project approval categories, principal investigator structures, and completion rates, identifying problems in university-level project management such as blocked development channels for young master's degree holders and high dependence of completion rates on achievement requirements. The study proposed the need to improve the construction mechanism for young research talent echelons [1]. Qin Zhixia pointed out that universities need to strengthen the whole-process supervision of research projects and build an integrated management platform to release the innovative value of research projects [4]. Based on the policy background of "delegation, regulation, and service," Zhang Hong constructed a four-stage performance evaluation system covering project establishment, implementation, acceptance, and tracking [5].

The educational value of horizontal research projects has gradually gained attention. Yu Jixian et al. argued that horizontal projects are often treated "discriminatorily" in university research evaluation systems, yet they play an irreplaceable role in the joint cultivation of professional degree postgraduates through school-enterprise cooperation [9]. Gan Lin et al. explored the dynamic accounting issues of horizontal projects from an accounting processing perspective, emphasizing their role as a main direction for universities to serve society [7].

2.2 Research on University Teacher Evaluation and Research Behavior

Research projects hold unique value in teacher evaluation. Zou Yongxing pointed out that due to their competitiveness, hierarchical nature, and strict peer review, research projects have become an important benchmark for teacher evaluation, but this has also led to the drawback of "research projectization," namely overemphasizing the number of approved projects while weakening research value [6-13]. Through questionnaire surveys, Yu Fangyan et al. found patterns in the impact of artificial intelligence technology on research projects of teachers at different levels: the pattern for professors was the same as that for teaching assistants, while the pattern for associate professors was the same as that for lecturers [2].

From the perspective of organized research, Shi Yueqi et al. proposed that university research needs to promote reforms in project management systems and evaluation systems to activate independent innovation vitality [8]. Jin Hua focused on the transformation of research achievements, emphasizing the strengthening of research project establishment mechanisms and motivation cultivation mechanisms [12].

2.3 Application of Data Mining in Research Management

Data mining technology provides new tools for research management decision-making. Taking Harbin Institute of Technology as an example, Liu Xing et al. analyzed the full lifecycle management of research project data and proposed improving the modernization level of research management through data governance [10]. Liu Donglin designed a personalized recommendation system for research resources based on collaborative filtering to meet the complex needs of teachers and students [11]. International scholars Febriana et al. combined team projects with flipped classrooms and validated their effectiveness in multidisciplinary writing teaching through the ADDIE framework [14]. Gabriela et al. studied the functional evolution of project management offices in university research centers and proposed a three-stage maturity model [15].

In summary, existing research mostly focuses on project management processes or teacher evaluation systems, with insufficient micro-level typological classification of teachers' research behavior. This study employs K-means clustering, taking a secondary college of a vocational undergraduate university as a case, to deeply analyze the internal structure of teachers' research behavior.

3. Research Design

3.1 Data Source and Preprocessing

3.1.1 Data Source

The data for this study originates from the statistical detailed list of scientific research projects at or above the municipal level undertaken by teachers of the School of Information Engineering, Hainan Vocational University of Science and Technology, from 2019 to 2025. A total of 90 project records were collected, encompassing information such as project title, principal investigator's name, educational background, professional title, project category, project number, and project level. Each record represents an approved research project. The seven-year time span comprehensively reflects the research activity characteristics of the faculty. The raw data includes instances where some teachers have multiple approved projects, providing a foundation for subsequent aggregation of features at the individual teacher level.

3.1.2 Data Cleaning

Data cleaning is a crucial step to ensure the reliability of analysis results. In the raw data, a small number of records had null values in the educational background or professional title fields. These records were excluded from teacher feature extraction due to missing key information. After cleaning, 86 valid project records were obtained. Simultaneously, the educational background and professional title fields were standardized by removing leading and trailing spaces to ensure consistent terminology within the same category.

3.1.3 Teacher Feature Extraction

Using the teacher's name as a unique identifier, multiple project records for each teacher were consolidated, and the following seven-dimensional features were extracted:

Highest Professional Title Rank: Professional titles were converted into ordinal numerical values. Based on the conventional hierarchy of university faculty titles, Professor (including other senior professional titles) was assigned a value of 4, Associate Professor a value of 3, Lecturer (including other mid-level titles) a value of 2, and Teaching Assistant (including unrated) a value of 1. The highest professional title appearing in a teacher's project records over the years was taken as their professional title rank, reflecting their current academic standing.

Highest Educational Background Rank: Doctorate was assigned a value of 3, Master's degree a value of 2, and Bachelor's degree a value of 1. The highest degree obtained was used to represent their academic training background.

Total Number of Projects: The cumulative number of projects undertaken by a teacher, reflecting their research activity level.

Quantities by Project Category: Based on the project category field, projects were classified into five main types:

Provincial-level Vertical Projects: Including Hainan Provincial Natural Science Foundation, Hainan Provincial Higher Education Scientific Research Projects, Hainan Provincial Higher Education Teaching Reform Projects, Hainan New Star Projects, Hainan Provincial Social Science Projects, etc.

Ministry of Education Collaborative Education Projects: Including the Ministry of Education's Higher Education Department Industry-Academia Cooperative Education Projects, Ministry of Education Industry-Academia Cooperative Education Projects, China University Industry-University-Research Innovation Fund Projects, Ministry of Education University Student Employment Education Projects, etc.

Horizontal Projects: Referring to collaborative projects with enterprises.

Municipal-level Projects: Such as Haikou City Philosophy and Social Science Planning Projects.

Others: Projects not clearly falling into the above categories, which were very few in this dataset and were temporarily not included in the main features.

After feature extraction, complete feature data for 40 teachers was obtained, forming the data feature matrix. The code for aggregating by teacher name is shown in Figure 1 Code for Aggregation by Teacher Name.

```
[1]: # ==================== 4. Aggregation by Teacher Name ====================
teacher_features = df_clean.groupby('Name').agg(
    max_title_rank=('Professional Title Rank', 'max'),
    max_edu_rank=('Education Level', 'max'),
    total_projects=('Project Name', 'count'),
    provincial_vertical=('Project Category', lambda x: (x == 'Provincial-level Vertical Projects').sum()),
    collab_edu=('Project Category', lambda x: (x == 'Collaborative Education Projects').sum()),
    horizontal=('Project Category', lambda x: (x == 'Horizontal Projects').sum()),
    municipal=('Project Category', lambda x: (x == 'Municipal-level Projects').sum()),
    other_projects=('Project Category', lambda x: (x == 'Others').sum())
).reset_index()

# Remove teachers with zero total projects
teacher_features = teacher_features[teacher_features['total_projects'] > 0]
print(f"Number of valid teachers: {len(teacher_features)}")

# ==================== 5. Prepare Clustering Features ====================
feature_cols = ['max_title_rank', 'max_edu_rank', 'total_projects',
                'provincial_vertical', 'collab_edu', 'horizontal', 'municipal']
X = teacher_features[feature_cols].values

# Standardization
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Figure 1: Code for Aggregation by Teacher Name.

3.2 K-means Clustering Method

K-means clustering is a classic partitioning-based clustering algorithm. Its core idea is to iteratively partition samples into K clusters, maximizing intra-cluster similarity and inter-cluster difference [16].

Since the professional title rank ranges from 1 to 4 and the total number of projects can be up to 6, the scales of these features differ. Using the raw data directly would allow features with larger magnitudes to dominate the clustering results. Therefore, Z-score standardization was applied: $z = (x - \mu)/\sigma$, making each feature have a mean of 0 and a standard deviation of 1, thus eliminating the influence of scale.

3.3 Method for Determining the Optimal K Value

In K-means clustering, the number of clusters, K, must be pre-specified. The choice of K directly affects the rationality and interpretability of the clustering results. This study comprehensively employs two methods, the Elbow Method and the Silhouette Coefficient, to determine the optimal K value.

3.3.1 Elbow Method

The Elbow Method is an intuitive technique for determining the optimal number of clusters by observing the trend of the Sum of Squared Errors (SSE) as a function of K [17]. The formula for SSE is given in equation (1):

$$\mathrm{SSE} = \sum_{k=1}^{K} \sum_{x \in C_k} \| x - \mu_k \|^2 \tag{1}$$

Here, C_K represents the k-th cluster, and μ_k is the centroid of that cluster. SSE measures the sum of squared distances of all samples to their respective cluster centroids, reflecting the compactness of the clustering results. As K increases, the samples are partitioned more finely, and SSE gradually decreases. When K is less than the true number of clusters, increasing K significantly reduces SSE; when K approaches or exceeds the true number of clusters, the rate of decrease in SSE slows down. The point where the rate of decrease sharply changes, forming an "elbow," is considered the optimal K value [17].

3.3.2 Silhouette Coefficient

The Silhouette Coefficient is an internal evaluation metric that assesses clustering effectiveness by combining intra-cluster cohesion and inter-cluster separation [16]. For a single sample i, its silhouette coefficient is calculated in two steps:

Calculate the average distance from sample i to all other samples within its own cluster, denoted a(i), which is called the intra-cluster dissimilarity. A smaller a(i) indicates that sample i is more tightly clustered.

Calculate the average distance from sample i to all samples in another cluster C_j, and take the minimum of these averages as b(i), i.e., b(i)=min_(j≠i) (average distance from sample i to cluster C_j). b(i) reflects the degree of separation between sample i and the nearest neighboring cluster. A larger b(i) suggests that sample i is less likely to belong to the other cluster.

The silhouette coefficient for sample i is defined as shown in equation (2):

$$\mathrm{s(i)} = \frac{\mathrm{b(i)} - \mathrm{a(i)}}{\max\{\mathrm{a(i)}, \mathrm{b(i)}\}} \tag{2}$$

The silhouette coefficient ranges from -1 to 1. When s(i) is close to 1, it indicates $a(i) \le b(i)$, meaning sample i is appropriately clustered. When s(i) is close to 0, the sample lies on the boundary between two clusters. A negative s(i) suggests the sample might have been assigned to the wrong cluster. The average silhouette coefficient for all samples provides an overall measure of the clustering quality.

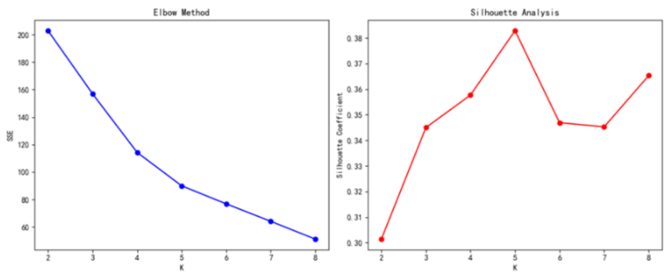
Based on its definition, this indicator comprehensively evaluates the effectiveness and rationality of clustering through the principles of minimizing internal distances and maximizing external distances. In practice, by calculating the average silhouette coefficient for different K values, the K value that maximizes the average silhouette coefficient is chosen as the optimal number of clusters.

3.4 Optimal K Value Determination Results

This study calculated the SSE and average silhouette coefficient for K values ranging from 2 to 8. The code section is shown in Figure 2 Determining the Optimal K Value, and the results are presented in Figure 3 SSE and Silhouette Coefficient for Different K Values.

```
[11]: # ==================== 6. Determine the Optimal K Value ====================
sse = []
sil_scores = []
K_range = range(2, 9)

for k in K_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(X_scaled)
    sse.append(kmeans.inertia_)
    sil_scores.append(silhouette_score(X_scaled, kmeans.labels_))

# Plot the elbow method graph
plt.figure(figsize=(12,5))
plt.subplot(1,2,1)
plt.plot(K_range, sse, 'bo-')
plt.xlabel('K')
plt.ylabel('SSE')
plt.title('Elbow Method')

plt.subplot(1,2,2)
plt.plot(K_range, sil_scores, 'ro-')
plt.xlabel('K')
plt.ylabel('Silhouette Coefficient')
plt.title('Silhouette Analysis')
plt.tight_layout()
plt.savefig('elbow.png', dpi=300)
plt.show()

# Select K with the maximum silhouette coefficient
best_k = K_range[np.argmax(sil_scores)]
print(f"Optimal K Value: {best_k} (Silhouette Coefficient={max(sil_scores):.3f})")

# Organize results into a DataFrame
table1 = pd.DataFrame({
    'K': list(K_range),
    'SSE': sse,
    'Silhouette Coefficient': sil_scores
})
print(table1.round(2))
```

Figure 2: Code for Determining Optimal K Value.




```
Optimal K Value: 5 (Silhouette Coefficient=0.383)
   K     SSE  Silhouette Coefficient
0  2  202.91                    0.30
1  3  156.91                    0.34
2  4  114.26                    0.36
3  5   89.91                    0.38
4  6   76.82                    0.35
5  7   64.04                    0.35
6  8   51.10                    0.37
```

Figure 3: SSE and Silhouette Coefficient for Different K Values.

As shown in Figure 3, SSE continuously decreases as K increases. When K increases from 4 to 5, SSE drops significantly from 114.26 to 89.91. However, for K > 5, the rate of decrease in SSE gradually slows down. According to the Elbow Method, K=5 is located at the inflection point where the rate of decrease changes from steep to gradual, which can be considered the "elbow point."

Regarding the silhouette coefficient, the average value slowly rises with increasing K, reaching 0.38 at K=5. Although there are slight increases for K=6 to 8, the magnitude is limited. Considering the simplicity and interpretability of the clustering results, along with the decreasing trend of SSE, K=5 was selected as the optimal number of clusters. This choice ensures good intra-cluster cohesion and inter-cluster separation while avoiding the interpretational difficulties associated with an excessive number of clusters.

3.5 Validation of Clustering Results

To verify the distinctiveness of the clustering results, one-way Analysis of Variance (ANOVA) was employed to test the significance of differences for each feature across different clusters. If the differences for each feature between clusters are statistically significant ($p < 0.05$), it indicates that the clustering effectively distinguishes different types of teacher groups. Additionally, Principal Component Analysis (PCA) was used to reduce the high-dimensional features to two dimensions for visualization, allowing observation of the distribution of each cluster in the reduced space, further validating the intuitive rationality of the clustering effect.

4. Analysis of Clustering Results

4.1 Cluster Centers and Cluster Characteristics

When K=5, the cluster centers for each cluster (restored to the original scale after standardization) and the sample distribution are shown in Figure 4 Mean Feature Values of Teachers by Cluster and Table 1 Mean Feature Values of Teachers by Cluster.

```
# ==================== 8. Output Clustering Results ====================
cluster_means = teacher_features.groupby('Cluster Label')[feature_cols].mean()
cluster_means['Number of Teacher'] = teacher_features['Cluster Label'].value_counts().sort_index()

print(cluster_means.round(2))


teacher_features.to_excel('Teacher Clustering Results.xlsx', index=False)
print("\nteacher_clustering_results.xlsx'")
```

Figure 4: Mean Feature Values of Teachers by Cluster.

Table 1: Mean Feature Values of Teachers by Cluster.

Serial Number	Cluster Type	Max Title Rank	Max Edu Rank	Total Projects	Provincial Vertical	Collaborative Education	Horizontal Projects	Municipal Projects	Number of Teachers
0	Teaching-Research Balanced Type	3.19	2	3.25	1.19	1.75	0	0	16
1	Novice Explorer Type	2.18	1.94	1.18	0.59	0.35	0	0	17
2	Local Service Type	3	2	1	0	0	0	1	2
3	Industry-Education Integration Type	4	2	2	0	0	2	0	1
4	Research Leader Type	3.75	3	3	2	0.5	0	0.5	4

4.2 Interpretation of Teacher Types

Based on the characteristics of each cluster, the five types of teacher groups are named as follows:

4.2.1 Cluster 0: Teaching-Research Balanced Type (16 individuals, 40.0%)

This group primarily consists of Associate Professors (mean value 3.19), all holding Master's degrees. They have a relatively high total number of projects (average 3.25 per person), balancing provincial-level vertical projects (1.19) and Ministry of Education collaborative education projects (1.75), with no horizontal or municipal-level projects. These teachers form the backbone of the college's research endeavors. They are capable of undertaking provincial-level vertical research (such as Natural Science Foundation projects or teaching reform projects) while actively participating in industry-academia collaborative education projects, reflecting the positioning of vocational undergraduate institutions that emphasize "equal emphasis on teaching and research, and integration of industry and education." They typically have established research directions and a certain level of research accumulation, but have not yet formed teams with doctoral degrees, and their capacity for horizontal expansion needs strengthening. This group aligns with the characteristics of the "main force of university-level projects" described in Yan Bingshu's study [1].

4.2.2 Cluster 1: Novice Explorer Type (17 individuals, 42.5%)

This group is mainly composed of Lecturers (mean professional title value 2.18), predominantly holding Master's degrees (1.94). They have a relatively low total number of projects (average 1.18 per person), primarily focusing on provincial-level vertical projects (0.59), with low participation in collaborative education projects (0.35) and no horizontal or municipal-level projects. These teachers are mostly young faculty members in the initial stages of their research careers, mainly relying on provincial-level vertical projects (especially teaching reform and scientific research projects) to accumulate experience. They face the "Matthew Effect" dilemma, with low success rates in project applications, and urgently need mentorship and project cultivation support. This finding highly coincides with Yan Bingshu's [1] observation of "blocked development channels for young teachers with Master's degrees" and also corroborates the findings of Yu Fangyan et al. [2] regarding the research patterns at the lecturer level.

4.2.3 Cluster 2: Local Service Type (2 individuals, 5.0%)

This group includes one Professor and one Lecturer, both holding Master's degrees. Each has a total of one project, all being municipal-level projects (Haikou City Philosophy and Social Science Planning Projects). These teachers focus on local economic and social development needs, undertaking municipal research projects, reflecting the function of serving the local community. Although the sample size is small, it indicates potential for the college in serving local decision-making consultation. This has some connection to the role of "technical experts" in the transformation of scientific research achievements emphasized by Jin Hua [12], although their output leans more towards policy consultation rather than technological transformation.

4.2.4 Cluster 3: Industry-Education Integration Type (1 individual, 2.5%)

This individual is a Professor with a Master's degree, having a total of two projects, both horizontal projects. This teacher primarily engages in horizontal research, directly addressing enterprise needs, embodying the characteristic of "applied research" in vocational undergraduate institutions. The educational value of their horizontal projects cannot be ignored [9], yet they are often subject to "discriminatory" treatment in traditional evaluation systems. This teacher, focusing on horizontal projects, may drive school-enterprise cooperation and cultivate students' practical abilities, aligning with the educational value of horizontal projects emphasized by Yu Jixian et al. [9].

4.2.5 Cluster 4: Research Leader Type (4 individuals, 10.0%)

This group primarily consists of Professors (mean value 3.75), all holding Doctoral degrees. They have a high total number of projects (3 items), mainly provincial-level vertical projects (2 items), with a small number of collaborative education and municipal-level projects. These teachers are the research leaders of the college. Possessing doctoral degrees and high professional titles, they undertake high-level vertical projects and produce stable output. They typically lead teams and are the core force for pursuing national-level projects and disciplinary development, conforming to the characteristics of high-level talent described by Zou Yongxing [6]. Additionally, teachers in Cluster 4 participate in a small number of municipal-level projects, possibly involving local service.

4.3 Visualization of Clustering Results

To intuitively display the clustering effect, Principal Component Analysis (PCA) was used to reduce the seven-dimensional features to two dimensions. The sample distribution plot is shown in Figure 5 (PCA Dimensionality Reduction Visualization of K-means Clustering). In the figure, each cluster shows clear separation in the reduced space: Cluster 0 (Teaching-Research Balanced Type) is concentrated in the center-left area; Cluster 1 (Novice Explorer Type) is more scattered but mainly located in the lower-left; Cluster 2 (Local Service Type) is situated in the upper-right; Cluster 3 (Industry-Education Integration Type) is isolated on the right; and Cluster 4 (Research Leader Type) is clustered in the lower-right. The boundaries between clusters are relatively clear, indicating that the selected features effectively distinguish different modes of teacher research behavior.

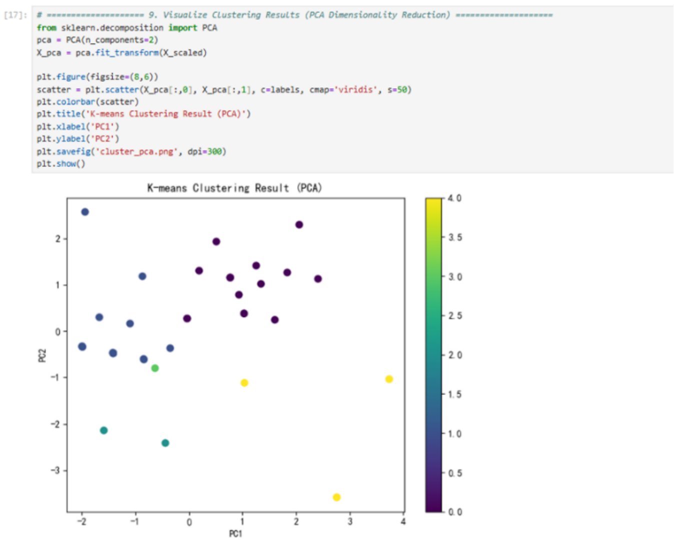


Figure 5: PCA Dimensionality Reduction Visualization of K-means Clustering.

5. Discussion

5.1 Exploration of Connections with Existing Research

The clustering results of this study corroborate several viewpoints from the literature:

First, the developmental difficulties faced by young teachers. Cluster 1 (Novice Explorer Type) accounts for 42.5% of all teachers, with a small number of projects primarily at lower-level vertical types. This aligns with Yan Bingshu's [1] finding regarding "blocked development channels for young teachers with Master's degrees." These teachers, mostly holding Master's degrees and with lower professional titles, are at a disadvantage in competing for research resources, easily falling into a vicious cycle of "difficulty in project approval—lack of accumulation—even greater difficulty in project approval."

Second, the unique value of horizontal projects. Although Cluster 3 (Industry-Education Integration Type) contains only one individual, it highlights the significance of horizontal projects in vocational undergraduate institutions. Yu Jixian et al. [9] pointed out the irreplaceable educational value of horizontal projects in school-enterprise joint cultivation. This study provides empirical support for this view: this teacher, focusing on horizontal projects, may involve students in real-world projects, thereby promoting the cultivation of practical abilities.

Third, the agglomeration effect of high-level talent. Cluster 4 (Research Leader Type), consisting mainly of doctoral-level professors, has high project output. This aligns with the resource agglomeration phenomenon under the background of "research projectization" analyzed by Zou Yongxing [6]. These teachers typically gain access to more research resources, forming a positive cycle, whereas teachers in Cluster 1 experience the opposite.

Fourth, the positioning emphasizing both teaching and research. Cluster 0 (Teaching-Research Balanced Type), the largest group, balances vertical projects and collaborative education, reflecting the characteristic of integrating research and teaching in vocational undergraduate institutions. This echoes the concept of multi-party collaboration in "organized research" proposed by Shi Yueqi et al. [8], namely that research should not only pursue academic innovation but also serve talent cultivation.

5.2 Implications for Research Management

Based on the clustering results, it is recommended that the college implement differentiated support strategies:

For the Teaching-Research Balanced Type (Cluster 0): Encourage them to strive for high-level vertical projects (such as National Natural Science Foundation projects) while deepening their involvement in collaborative education projects. Leverage their experience to mentor young teachers through guidance and support. Additionally, increase the weight of teaching-research projects in professional title evaluations to reflect the mutual reinforcement of teaching and research.

For the Novice Explorer Type (Cluster 1): Implement a "research mentorship system," pairing them with experienced core teachers for guidance. Establish college-level youth startup funds to lower the threshold for project applications. Strengthen research methodology training to improve the quality of project proposals. Drawing on the research of Yu Fangyan et al. [2], conduct targeted training for teachers at different levels.

For the Local Service Type (Cluster 2): Build platforms to connect with local needs, encouraging them to transform municipal project outcomes into policy consultation or horizontal collaborations. Establish evaluation mechanisms for local service achievements, incorporating policy recommendations, planning proposals, etc., into research performance assessments.

For the Industry-Education Integration Type (Cluster 3): Increase the weight of horizontal projects in professional title evaluations and performance assessments to eliminate "discriminatory" evaluation [9]. Encourage them to transform horizontal outcomes into teaching cases, achieving the integration of science and education.

For the Research Leader Type (Cluster 4): Support them in forming interdisciplinary teams to pursue major national projects. Encourage collaboration with industry to promote the transformation of achievements. Simultaneously, require them to undertake the task of cultivating young teachers to build a talent development pipeline.

Furthermore, the college should strengthen the connection between projects at different levels, constructing a "university-level → provincial-level → national-level" cultivation chain [1], and establish a diversified outcome evaluation system [13] to avoid an over-reliance on project quantity as the sole metric.

5.3 Research Limitations

This study has the following limitations: First, the sample size is relatively small (40 teachers), with extremely few individuals in Cluster 2 and Cluster 3, so conclusions should be generalized with caution. Second, the feature selection is limited to project quantity and type, without considering potential factors such as teacher age, research direction, or publication output. Third, the data originates from only one college, so extrapolating the findings requires prudence. Future research could expand the sample scope, introduce more variables, conduct comparative studies across institutions, and attempt to use methods like deep learning to uncover deeper-level patterns.

6. Conclusion

Based on the data of municipal-level and above research projects undertaken by teachers at the School of Information Engineering, Hainan Vocational University of Science and Technology, from 2019 to 2025, this study employed K-means clustering to classify 40 teachers into five types: Teaching-Research Balanced Type, Novice Explorer Type, Local Service Type, Industry-Education Integration Type, and Research Leader Type. The results reveal significant differences among these teacher types in terms of professional title structure, educational background, project quantity, and project type preferences. The clustering outcomes provide data support for precision research management at the college level and corroborate existing research findings, such as Yan Bingshu's [1] observation regarding blocked development channels for young teachers and Yu Jixian et al.'s [9] emphasis on the educational value of horizontal projects.

As the construction of Hainan Free Trade Port advances and the 15th Five-Year Plan is fully implemented, vocational undergraduate universities will assume a more significant mission in the integrated development of education, technology, and talent. The School of Information Engineering should leverage the differentiated advantages of the five teacher groups, precisely align with Hainan Province's future industrial demands in digital economy, artificial intelligence, and other fields, deepen industry-education integration, and encourage teachers to derive research topics from industrial needs and apply their achievements to serve local development. Simultaneously, it is necessary to further refine the categorized evaluation system: emphasizing academic innovation for the Research Leader Type, strengthening the orientation toward achievement transformation for the Industry-Education Integration Type, and implementing developmental evaluations for new teachers, thereby fostering a research ecosystem conducive to the growth of all teacher types. Future research could expand the sample scope, include more institutions for comparative analysis, and introduce longitudinal tracking data to explore the evolutionary patterns of teachers' research behavior, providing a more solid empirical basis for reforming research management in vocational undergraduate universities.

Acknowledgments

This paper was supported by the Hainan University of Science and Technology School-level Scientific Research Funding Project + Path Research on Big Data Technology Empowering Rural Revitalization in the Context of Hainan Free Trade Port: HKKY2024-47.

This research has received support from a provincial-level project of the 2025 College Students' Innovation and Entrepreneurship Program at Hainan Technology and Business University: Design and Implementation of an Intelligent Monitoring and Management System for Used Clothing Recycling Bins Based on STM32 and WeChat Mini Program.

References

[1] Yan Bingshu. Research on the Improvement Path of University-Level Scientific Research Project Management—Taking a Open University as an Example. Huazhang, 2026, (02): 132-134.

[2] Yu Fangyan, Han Xiangdong. Statistical Research on the Professional Development of University Teachers under the Background of Artificial Intelligence // Jiangxi Provincial Association for the Promotion of Culture and Arts, Nanchang Institute of Technology. Proceedings of the First Symposium on Innovation and Development of New Liberal Arts Empowered by Artificial Intelligence. 2026: 667-670.

[3] Wan Min, Chen Mou, Shao Shuyi. Reform and Practice of Experimental Teaching in Automation Majors in Universities Driven by Scientific Research Projects. Journal of Higher Education, 2026, 12(01): 123-126.

[4] Qin Zhixia. Enhancing the Efficiency of Scientific Research Management to Consolidate the Foundation of University Innovation. China Reform Daily, 2025-12-03 (004).

[5] Zhang Hong. Research on the Performance Evaluation System of University Scientific Research Projects under the Background of "Delegation, Regulation, and Service". Journal of Jilin Radio and TV University, 2025, (06): 122-124.

[6] Zou Yongxing. The "Projectification of Scientific Research" in University Teacher Evaluation. Journal of Educational Science of Hunan Normal University, 2025, 24(03): 101-110.

[7] Gan Lin, Zhang Yu. Discussion on the "Dynamic" Accounting Treatment of Input Tax Deduction for Horizontal Scientific Research Projects in Universities. Communications of Finance and Accounting, 2024, (19): 171-176.

[8] Shi Yueqi, Zhou Haitao. Characteristics, Models, and Strategies of Organized Research in Universities. Journal of Higher Education Research, 2023, 44(02): 30-36.

[9] Yu Jixian, Lin Chengyan, Li Zhaomin. The Educational Value of Horizontal Scientific Research Projects in Universities—Based on the Perspective of Professional Degree Postgraduate Training in Engineering. Academic Degrees & Graduate Education, 2022, (03): 51-56.

[10] Liu Xing, Han Tingting, Wang Shengjin, et al. Challenges and Countermeasures for High-Quality Data Management of Scientific Research Projects in Chinese Universities—Taking Harbin Institute of Technology as an Example. R&D Management, 2021, 33(05): 161-167.

[11] Liu Donglin. Personalized Integration and Recommendation of Scientific Research Resources in Universities [J]. Journal of Chongqing University, 2022, 45(07): 122-130.

[12] Jin Hua. Construction of the Realization Mechanism for the Transformation of Scientific Research Achievements in Universities. China University Science & Technology, 2019, (12): 87-89.

[13] Zou Yongxing. The "Projectification of Scientific Research" in University Teacher Evaluation. Journal of Educational Science of Hunan Normal University, 2025, 24(03): 101-110. (Reprint)

[14] Febriana N, Ermanto, Abdurahman. Developing Team-Based Project Model Combined with Flipped Classroom for Scientific Writing in Higher Education. International Journal of Information and Education Technology, 2025, 15(10).

[15] Gabriela F, Hugo S, Anabela T, et al. Role of the Project Management Office in University Research Centres. Sustainability, 2021, 13(21):12284.

[16] Zhang Dongmei. Research on Hierarchical Clustering Algorithm Based on Silhouette Coefficient. Yanshan University, 2010.

[17] Sun Jing. Research on Weighted K-means Clustering Algorithm Based on Python. China New Technologies and New Products, 2024, (19): 1-4. DOI: 10.13612/j.cnki.cntp.2024.19.005.