Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 2, August 2025, pp. 1310~1325
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i2.pp1310-1325

Analyzing and clustering students admission data in Yala Rajabhat University Thailand

Thanakorn Pamutha[1], Wanchana Promthong[2], Sofwan Palawan[2]
[1]Faculty of Science Technology and Agriculture, Yala Rajabhat University (YRU), Yala, Thailand
[2]Academic Resources and Information Technology Center, Yala Rajabhat University (YRU), Yala, Thailand

Article Info

Article history:

Received Aug 24, 2024
Revised Mar 26, 2025
Accepted Jul 2, 2025

Keywords:

Admission data analysis
Cluster analysis techniques
Educational data mining
Student clustering
Student segmentation

ABSTRACT

This research explores the use of clustering techniques to analyze student admission data at Yala Rajabhat University, Thailand, aiming to enhance recruitment strategies and understand student profiles. Employing K-means, Hierarchical Clustering, and Density-based spatial clustering of applications with noise (DBSCAN), the study groups admission data based on factors like educational institution, geographic location, and program chosen. The methodology incorporates normalization and principal component analysis (PCA) to ensure data quality, while the Elbow Method determines the optimal number of clusters for effective data segmentation. The davies-bouldin index (DBI) evaluates the clustering configurations, ensuring that clusters are well-separated and cohesive. The results reveal distinct student profiles that can inform targeted marketing and improve recruitment strategies. This study not only provides strategic insights into student recruitment but also contributes to the literature on the use of data science in educational settings, highlighting the transformative impact of advanced analytics on institutional effectiveness. The research emphasizes the importance of data-driven approaches in adapting to the changing dynamics of student admissions and the competitive landscape of higher education.

Corresponding Author:

Thanakorn Pamutha
Faculty of Science Technology and Agriculture, Yala Rajabhat University (YRU)
133 Thetsaban 3 Road, Tambol Sateng, Amphoe Mueang, Yala Province, 95000, Thailand
Email: thanakorn.p@yru.ac.th

1. INTRODUCTION

Higher education institutions in Thailand face a critical challenge due to declining birth rates and increased competition among universities [1], [2]. Traditional methods of student recruitment, such as mass marketing, career fairs, and school visits, are proving ineffective in attracting the right students. This situation necessitates a data-driven approach to analyzing student admission patterns, enabling universities to optimize recruitment strategies and resource allocation.

One key method for improving university admissions is clustering analysis, which can segment students based on various attributes such as geographical background, academic performance, and institutional preference. Recent literature supports using methods like K-means clustering to segment student populations based on various demographic and academic variables, allowing universities to identify distinct groups within their applicant pools for targeted engagement [3]-[15].

Advanced clustering techniques, including hierarchical clustering and density-based spatial clustering of applications with noise (DBSCAN), have effectively handled diverse student admission datasets. Unlike K-means, which are sensitive to cluster shape and size, DBSCAN is particularly useful for detecting arbitrary-shaped clusters and outliers, making it valuable for complex student data segmentation.

Feature selection is vital in optimizing clustering performance, especially in student admission analysis. Recent studies have proposed various feature selection techniques to enhance clustering outcomes, ensuring that only the most relevant attributes are retained for analysis.

Moreover, effective data preprocessing techniques such as data cleaning, normalization, and dimensionality reduction are critical in improving the accuracy and reliability of clustering results. Principal component analysis (PCA) is widely applied as a dimensionality reduction technique, helping streamline high-dimensional student admission data while preserving key information. These preparatory steps ensure that clustering models produce actionable and interpretable insights, supporting data-driven decision-making in higher education institutions.

Following a thorough review of relevant literature and research studies, several investigations have explored the application of clustering techniques alongside various datasets to enhance student admission analysis. These studies have demonstrated the effectiveness of clustering methods in segmenting student populations based on academic and demographic attributes. Table 1 [16]-[30] (in Appendix) provides a comprehensive summary of these findings, highlighting key methodologies and results from prior research.

Several studies have explored data mining techniques in education, particularly in student performance prediction and dropout analysis. However, gaps remain in applying clustering techniques for admission data analysis, as prior studies predominantly focus on predicting student academic performance rather than optimizing student recruitment strategies. Additionally, few research efforts have compared multiple clustering techniques such as K-means, DBSCAN, and Hierarchical Clustering within real student admission datasets. Moreover, the lack of proper validation metrics, such as the davies-bouldin index (DBI), limits the ability to determine the most effective clustering approach for student segmentation. Given these gaps, there is a pressing need to investigate advanced clustering approaches for student admission data to enhance decision-making processes in higher education institutions.

This research makes the following key contributions to the field of student admission data analysis and clustering:

- Develop a novel clustering framework integrating PCA for feature selection and DBI for validation to improve clustering accuracy and interpretability.
- Conducts a comparative evaluation of clustering techniques, including K-means, DBSCAN, and Hierarchical Clustering.
- Applies clustering techniques to admission data from Yala Rajabhat University to provide practical insights for recruitment and academic planning.
- Supports data-driven decision-making in university admissions by facilitating targeted recruitment strategies based on well-defined student clusters.

The remainder of this paper is organized as follows: Section 2: Methodology – Describes data collection, preprocessing, and feature selection using PCA. It explains clustering techniques, including K-means, DBSCAN, and Hierarchical Clustering, and introduces the DBI for clustering validation. Section 3: Results and Analysis – Presents clustering results, evaluates performance using DBI scores, and visualizes clustering outcomes. Section 4: Discussion – Interprets findings, assesses clustering performance, and explores its practical applications in student admissions. Section 5: Conclusion and Future Work – Summarizes key research findings, discusses study limitations, and proposes directions for future research.

2. METHOD

This section describes the research methodology used to analyze and cluster student admission data at Yala Rajabhat University. The data was preprocessed, clustered, and evaluated for optimal segmentation. The methodology follows a systematic approach, as illustrated in Figure 1.

The following application of the data mining clustering model on analyzing and clustering students who choose to Study at Yala Rajabhat University as follows:

2.1. Data collection

Student admission history data was collected from the Educational Services Division, Office of the President, Yala Rajabhat University. The dataset comprises records from the academic years 2019-2023, totaling 13,435 data entries. Key attributes include sex, religion, hometown province, school, school plan, grade point average (GPA), program, and faculty. To ensure privacy, all identifying information was anonymized. The data was extracted from institutional databases to maintain accuracy and integrity.

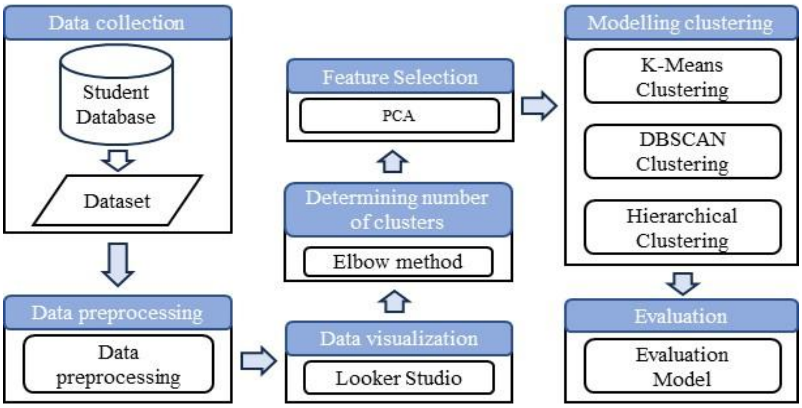

Figure 1. Research methodology

2.2. Data preprocessing

Data preprocessing ensures the dataset is clean, structured, and formatted correctly for clustering analysis. This step involves nominal data encoding, cleaning, and standardization to improve processing efficiency and consistency.

2.2.1. Nominal data encoding

Nominal data encoding is essential for converting categorical variables into numerical formats suitable for clustering algorithms. This study's categorical variables, such as school type, province, and study plan, were transformed using numeric codes. This conversion allowed algorithms to process the data efficiently without misinterpretation due to categorical values.

2.2.2. Data cleaning

Data cleaning involves removing duplicate records, handling missing values, and resolving inconsistencies. Attributes such as sex and religion were omitted to reduce noise and enhance the relevance of clustering. The dataset was further examined for outliers and erroneous entries, which were either corrected or removed based on established preprocessing criteria. The refined dataset is summarized in Table 2, showing encoded and cleaned attributes.

Table 2. Nominal data encoding and cleaned dataset

No	SchCode	SchProCode	SchTypeCode	SchPlanCode	GpaxCode	Program	Facultycode
1	1	95	11	1	3	2	1
2	1	95	11	2	4	2	1
3	3	91	1	2	3	2	1
4	4	94	1	2	3	2	1
5	5	96	11	2	3	2	1
6	6	94	11	2	4	2	1
7	7	95	11	2	3	2	1
8	8	91	11	2	4	2	1

2.2.3. Data standardization

Standardize the data to ensure that all numerical variables are brought to a uniform scale, which may also involve encoding categorical variables. This standardization process entails transforming the data by subtracting the mean and dividing it by the standard deviation. This step is crucial to mitigate biases that could arise when variables with larger scales disproportionately affect the outcomes of the clustering algorithm.

2.2.4. Determining the number of clusters

Identifying the optimal number of clusters for K-means and hierarchical clustering algorithms using the Elbow Method [22], [31]. The value of k Identifies the optimal number of clusters that best represent the underlying patterns in the data [22]. Implement methods like the Elbow Method, which involves plotting the sum of squared distances from each point to its assigned center and choosing the point where improvements become marginal.

2.3. Feature selection

Feature selection enhances clustering accuracy by reducing dimensionality and retaining only the most relevant attributes [25]. PCA was applied to identify the key features contributing to variance in the dataset [26]. This widely used dimensionality reduction technique transforms high-dimensional data into a lower-dimensional space while preserving as much variance as possible [27]. In the context of student admission data analysis, PCA helps select the most important features contributing to clustering, ensuring efficient data processing and improved cluster quality [28]. By applying PCA, this study effectively identified the most important student attributes contributing to meaningful clustering. The method reduced data dimensionality, improved clustering accuracy, and facilitated better segmentation of students for targeted admission strategies [28].

2.4. Clustering

Three clustering techniques were applied to segment the student admission data: K-means, DBSCAN, and Hierarchical Clustering. Each algorithm provides distinct advantages, allowing for comprehensive analysis:

2.4.1. K-means clustering

K-means clustering is one of the most widely used unsupervised machine learning algorithms for partitioning a dataset into K distinct, non-overlapping clusters. It is particularly effective for well-separated and spherical clusters, making it a popular choice in educational data mining, including student admission analysis. The algorithm minimizes intra-cluster variance while maximizing inter-cluster separation, ensuring that data points within a cluster are more similar than those in different clusters [16], [23].

2.4.2. DBSCAN

DBSCAN is a density-based clustering algorithm that groups together data points that are closely packed while identifying points that lie in low-density regions as outliers [23]. Unlike K-means, which require specifying the number of clusters beforehand, DBSCAN automatically determines the number of clusters based on data distribution [32]. This makes it especially effective for datasets with clusters of arbitrary shape and varying densities, including student admission data, where student groups may not have clear, spherical distributions [24].

2.4.3. Hierarchical clustering

Hierarchical Clustering is a powerful unsupervised machine learning technique used to group similar data points into a tree-like structure, known as a dendrogram [23]. Unlike K-means and DBSCAN, which require predefined parameters for the number of clusters, Hierarchical Clustering forms a hierarchy of nested clusters, allowing flexibility in cluster selection at different levels. This method is particularly useful in student admission analysis, as it helps universities identify relationships between student groups based on academic backgrounds, geographical regions, and program choices. By visualizing the clustering process as a tree, institutions can explore student similarities at various levels of granularity [24].

2.5. Model evaluation

Model evaluation using the DBI measures how similar an object is to its own cluster compared to other clusters, ensuring a good clustering configuration. The DBI evaluates clustering quality by assessing both the compactness and separation of clusters, measuring how well they are distinct from each other and compact within themselves [29], [32]. It calculates the average similarity ratio between each cluster and its most similar neighboring cluster, considering both the intra-cluster distance (the average distance between points within the same cluster) and the inter-cluster distance (the distance between the centroids of different clusters). A lower DBI value indicates better clustering performance, characterized by compact clusters that are well-separated from each other, while a higher DBI value suggests poorer clustering with overlapping and less distinct clusters [29], [30]. This index is particularly useful for comparing different clustering results and determining the optimal number of clusters within a dataset, making it a valuable tool for analyzing and validating clustering algorithms.

2.6. Tool and data visualization

The Python programming language was used extensively for data preprocessing, clustering, and analysis due to its powerful libraries such as Pandas for data manipulation, Scikit-learn for machine learning algorithms, and Matplotlib for data visualization. Clustering algorithms, including K-means, Hierarchical Clustering, and DBSCAN, were implemented in Python to segment the student data into meaningful clusters. Looker Studio was utilized to create comprehensive visual representations of both the descriptive analysis and the clustering results.

The analysis included data cleaning, standardization, and PCA to prepare data for clustering. We employed K-means, Hierarchical, and DBSCAN algorithms, evaluating clustering quality with the DBI. This methodological framework ensures a robust analysis and a clear understanding of data patterns and facilitates precise marketing and strategic planning.

3. RESULTS

The thorough analysis and clustering of student admission data at Yala Rajabhat University have produced key insights that are set to enhance the university's recruitment strategies and student engagement. This section details the results and discussion of using various clustering algorithms—K-means, Hierarchical Clustering, and DBSCAN—to categorize admission data by educational institution, geographic location, and chosen programs. These methods have successfully highlighted various student profiles, which are clearly illustrated through figures and tables, showcasing the effectiveness of the clustering process.

3.1. Descriptive analysis

This section provides a general overview of the student population at Yala Rajabhat University from the academic years 2019 to 2023, which includes a total of 13,435 students. The breakdown by faculty is detailed in Table 3. Having established a clear understanding of the overall student demographics and preferences, we now focus on a more detailed examination through clustering analysis. By applying various clustering algorithms, we aim to uncover distinct student profiles that can further inform strategic planning and recruitment efforts.

According to Table 1, the Faculty of Humanities and Social Sciences has the highest enrollment, followed by the Faculty of Management Sciences, the Faculty of Science, Technology, and Agriculture, and the Faculty of Education. Most students originate from the southern region, particularly the southern border provinces. Yala Province has the highest number of students, followed by Pattani, Narathiwat, Songkhla, Satun, and other provinces, as depicted in Table 4. The majority of students previously attended Thamwittayamulniti School in Yala Province, followed by Darussalam School in Narathiwat Province, and other noted institutions listed in Table 5. Currently, Yala Rajabhat University offers 54 different programs. The most popular program students choose is the Bachelor of Political Science, followed by the Bachelor of Laws, Bachelor of Accounting, and Management program, as detailed in Table 6.

Table 3. Number of students by faculty (2019–2023)

No	Faculty	Number	Percentage
1	Faculty of Humanities and Social Sciences	5,213	38.80
2	Faculty of Management Sciences	3,669	27.31
3	Faculty of Science, Technology and Agriculture	2,301	17.13
4	Faculty of Education	2,252	16.76

Table 4. Students' school provinces (2019–2023)

No	Province	Number	Percentage
1	Yala	,4996	37.19
2	Pattani	,3785	28.17
3	Narathiwat	27,54	20.50
4	Songkla	630	4.69
5	Stul	544	4.05
6	Other	726	5.40

Table 5. Top 10 previous educational institutions of students (2019–2023)

No	School name	Province	Number	Percentage
1	Dhammavidya Mulniti School	Yala	785	5.84
2	Darussalam School	Narathiwat	501	3.73
3	Yala Community College	Yala	478	3.56
4	Narathiwat Community College	Narathiwat	445	3.31
5	Drun Satsana Witthaya	Pattani	429	3.19
6	Pattani Community College	Pattani	356	2.65
7	Yala Vocational College	Yala	280	2.08
8	Attarkiyah Islamiah School	Satul	264	1.97
9	Yala Patdung Pracha Vocational College	Yala	258	1.92
10	Phatthana Witthaya	Yala	225	1.67

Table 6. Top 10 most popular programs (2019–2023)

No	Program	Number	Percentage
1	Bachelor of Political Science	1,160	8.63
2	Bachelor of Laws	1,134	8.44
3	Bachelor of Accounting	1,027	7.64
4	Management	883	6.57
5	Early Childhood Care and Development	819	6.10
6	Business Computer and Digital Technology	683	5.08
7	Community Development Innovation	486	3.62
8	English Language	440	3.28
9	Early Childhood Education	350	2.61
10	Public Administration and Law	338	2.52

The initial descriptive analysis provides an overview of the student demographics and preferences at Yala Rajabhat University (Tables 3-6). Most students originate from the southern region, particularly Yala, Pattani, and Narathiwat provinces. Having established a clear understanding of the overall student demographics and preferences, we now turn our focus to a more detailed examination through clustering analysis. By applying various clustering algorithms, we aim to uncover distinct student profiles that can further inform strategic planning and recruitment efforts.

3.2. Clustering results

The following section presents the results obtained from applying clustering algorithms to the student admission data. This analysis aims to identify distinct groups within the student population based on factors such as educational institution, geographic location, and chosen programs.

3.2.1. Clustering performance evaluation

Determining the optimal number of clusters for K-means and hierarchical clustering algorithms using the Elbow Method. The Elbow method was used to determine the most appropriate number of clusters (k). The value of k=6 Figure 2 shows the result of the elbow method.

Table 7 presents the DBI for both K-means and hierarchical clustering algorithms using 6 clusters. Additionally, it specifies the 'eps' and 'minds' parameters for the DBSCAN algorithm as 0.2 and 300, respectively.

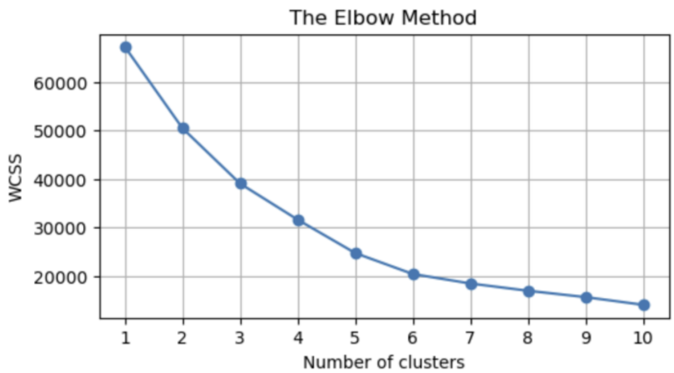

Figure 2. Elbow method to determine the most appropriate number of clusters

Table 7. DBI for clustering quality

PCA	db_index of K-means (K=6)	db_index of Hierarchical (K=6)	db_index of DBSCAN (eps=0.2, min_samples=300)
PCA2	0.7468	0.8165	1.9797
PCA3	0.8933	1.1621	2.1455
PCA4	0.8805	1.0347	1.3915
PCA5	1.0838	1.1928	1.0503

Table 7 compares the DBI for K-means and hierarchical clustering algorithms, both using 6 clusters, and specifies the 'eps' and 'minds' parameters for DBSCAN as 0.2 and 300, respectively. The table shows DBI values across different principal component analyses (PCA2, PCA3, PCA4, and PCA5), highlighting the clustering quality for each algorithm. The results indicate that K-means generally have lower DBI values than hierarchical and DBSCAN clustering, suggesting varying clustering effectiveness depending on the algorithm and the number of principal components used. Figures 3-5 displays scatter plots visualizing the clustering results obtained from K-means, hierarchical, and DBSCAN clustering algorithms.

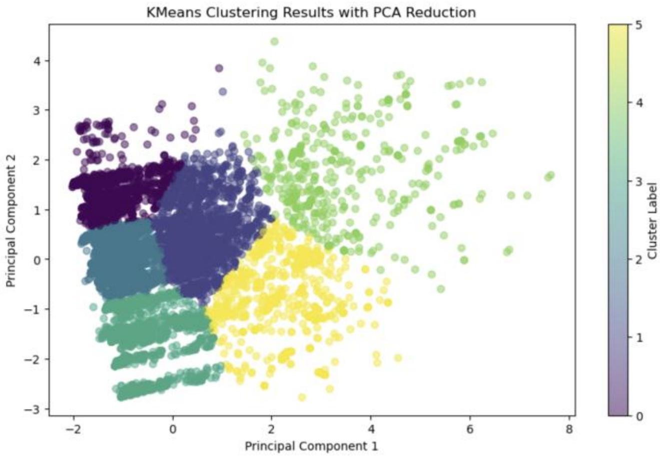

Figure 3. K-means clustering

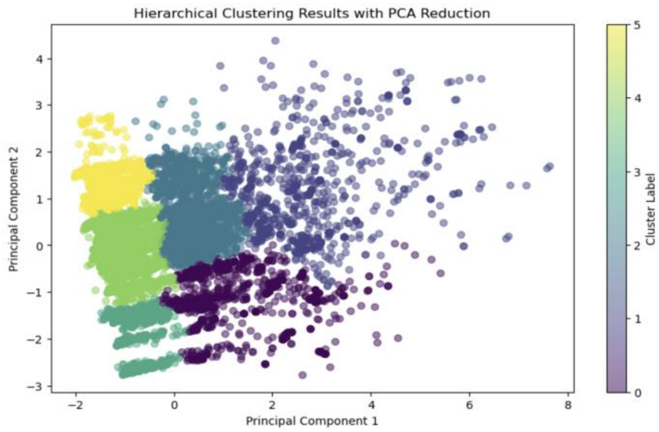

Figure 4. Hierarchical clustering

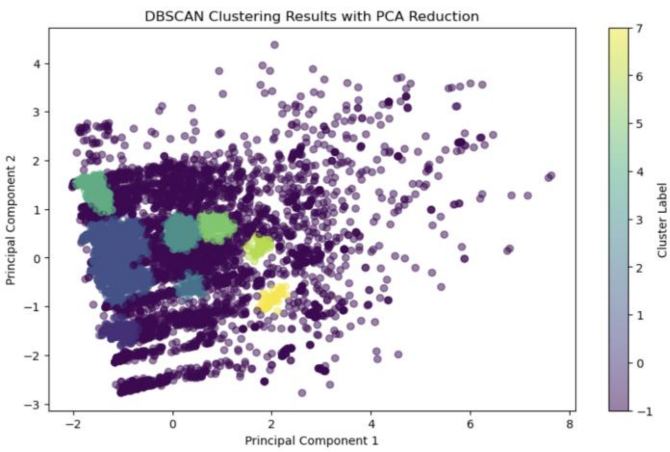

Figure 5. DBSCAN clustering

K-means Clustering creates distinct and well-separated clusters, effectively organizing the data into uniform groups. This indicates that K-means is highly suitable for this dataset, particularly beneficial for data where clear, defined groupings are needed, and the clusters are generally spherical and consistent in size. Hierarchical Clustering offers a more detailed perspective of the data structure, with clusters that blend gradually, providing a deeper understanding at various levels of detail. This approach could be beneficial for deciphering complex relationships within the student population, though it may not delineate clusters as sharply as K-means. DBSCAN successfully detects dense clusters and outliers, which is ideal for data featuring irregularly shaped clusters or different densities. However, the appearance of noise and blurred cluster boundaries may hinder DBSCAN's effectiveness in cases where precise segmentation is necessary or if the dataset lacks notable density differences.

The clustering analysis identifies distinct groups within the student population, each defined by unique demographic and academic characteristics. The results indicate that K-means clustering is highly effective for this dataset, providing well-structured and meaningful student groupings. The following section presents a detailed analysis of the K-means clustering results.

3.2.2. K-means clustering results

Analyzing student admission data from Yala Rajabhat University from 2019 to 2023 from over 13,435 students, the study highlights the distribution of students across various faculties and their geographical origins. The findings indicate that most students enroll in faculties such as Humanities and Social Sciences, Management Sciences, Science, Technology and Agriculture, and Education. The regional analysis reveals that most students come from the southern region of Thailand, particularly from the provinces of Yala, Pattani, and Narathiwat. This regional preference emphasizes the local demographic's affinity for Yala Rajabhat University, potentially driven by proximity and cultural familiarity. Furthermore, the popularity of specific programs, such as Political Science, Law, and Accounting, highlights targeted academic interests within the student body. These insights are crucial for the university’s strategic planning, enabling tailored recruitment strategies and optimized resource allocation. By aligning academic offerings with student preferences, Yala Rajabhat University can enhance educational outcomes and student satisfaction, thereby reinforcing its position in the competitive educational landscape.

This analysis of Clustering Students Admission Data at Yala Rajabhat University Thailand chose the K-means algorithm due to its highest clustering performance. It was found that it could divide the students into 6 groups as follows:

- Cluster 0: Has 1,629 students, accounting 12.10% from schools under the Office of the Private Education Commission located in Yala Province, with 1,453 students, making up 89.20%, Yala Province with 736 students, accounting for 45.18%, Pattani Province with 471 students, representing 28.91%, Narathiwat Province with 310 students, representing 19.03%, and Satul Province with 91 students, comprising 3.44%. The students have completed a study plan in Science and Mathematics with an average grade level of Medium (1.6 - 2.4). The top five programs chosen at Yala Rajabhat University are Bachelor of Public Administration, with 1,160 students, accounting for 8.63%; Bachelor of Laws, with 1,134 students, making up 8.44%; Bachelor of Accountancy, with 1,027 students, representing 7.64%, Management again with 883 students, accounting for 6.57%, and Early Childhood Care and Development with 819 students, comprising 6.10%.
- Cluster 1: Contains 3,305 students, accounting for 24.60% from various educational institutions, as shown in Table 8. These educational institutions are in various provinces, as shown in Table 9.

Most students have completed the Vocational Certificate program, with 1,270 students accounting for 38.43%, and the Science and Mathematics plan, 1,021 students, accounting for 30%, with an average grade level of Good (2.5 - 3.4). The top five programs chosen at Yala Rajabhat University are shown in Table 10.

Table 8. Previous educational institutions for student cluster 1

No	Educational institutions	Number	Percent
1	Secondary Educational Service Area	1,255	37.97
2	Vocational College	625	18.91
3	Office of the Private Education Commission	399	12.07
4	Industrial Community Education College	354	10.71
5	Technical College	252	7.62
6	Community College	183	5.54
7	Department of Learning Encouragement	137	4.15
8	Other	100	3.02
	Total	3,305	100

Table 9. Provinces of educational institutions in cluster 1

No	Educational institutions	Number	Percent
1	Yala	1,301	39.36
2	Pattani	934	28.26
3	Narathiwat	724	21.91
4	Songkla	158	4.78
5	Satul	47	1.42
6	Other	141	4.27
	Total	3,305	100

Table 10. Top five programs in cluster 1

No	Educational institutions	Number	Percent
1	Bachelor of Accounting	584	17.67
2	Business Computing and Digital Technology	463	14.01
3	Management	447	13.53
4	Bachelor of Laws	373	11.29
5	Bachelor of Public Administration	302	9.14

Located in Yala Province, with 1,453 students, making up 89.20%, Yala Province with 736 students, accounting for 45.18%, Pattani Province with 471 students, representing 28.91%, Narathiwat Province with 310 students, representing 19.03%, and Satul Province with 91 students, comprising 3.44%. The students have completed a study plan in Science and Mathematics with an average grade level of Medium (1.6 - 2.4). The top five programs chosen at Yala Rajabhat University are Bachelor of Public Administration, with 1,160 students, accounting for 8.63%; Bachelor of Laws, with 1,134 students, making up 8.44%; Bachelor of Accountancy, with 1,027 students, representing 7.64%, Management again with 883 students, accounting for 6.57%, and Early Childhood Care and Development with 819 students, comprising 6.10%.

– Cluster 2: Has 3,921 students, accounting for 29.20% of schools under the Office of the Private Education Commission, with 3,893 students making up 99.29%, Secondary Educational Service Area with 24 Students representing 4.61%. These educational institutions are in various provinces, as shown in Table 11.

Table 11. Provinces of educational institutions in cluster 2

No	Educational institutions	Number	Percent
1	Yala	1,579	40.27
2	Pattani	1,204	30.71
3	Narathiwat	831	21.19
4	Satul	206	5.25
5	Songkla	82	2.09
6	Other	19	0.49
	Total	3,921	100

Most students have completed the Science and Mathematics plan, with 3,225 students, accounting for 82.25%; the Arts and Languages plan, 405 students, accounting for 10.33% and the Arts and Mathematics plan, 288 students, accounting for 7.35% with an average grade level of Good (2.5 - 3.4). The top five programs chosen at Yala Rajabhat University as shown in Table 12.

– Cluster 3: Contains 2,094 students, accounting for 15.60% from various educational institutions, as shown in Table 13.

Table 12. Top five programs in cluster 2

No	Educational institutions	Number	Percent
1	Bachelor of Public Administration	279	7.12
2	Accounting	262	6.68
3	English	237	6.04
4	Bachelor of Laws	200	5.10
5	Public Health	170	4.34

Table 13. Previous educational institutions for student cluster 3

No	Educational institutions	Number	Percent
1	Office of the Private Education Commission	1,818	86.82
2	Secondary Educational Service Area	240	11.46
3	Other	37	1.72
	Total	2,094	100

These educational institutions are in various provinces, as shown in Table 14. Most students have completed the Science and Mathematics plan, with 1,870 students, accounting for 89.30%, and the Arts and Languages plan, with 127 students, accounting for 6.06% with an average grade level of Best (3.5 – 4.0). The top five programs chosen at Yala Rajabhat University as shown in Table 15.

Table 14. Provinces of educational institutions in cluster 3

No	Educational institutions	Number	Percent
1	Pattani	739	35.29
2	Yala	673	32.14
3	Narathiwat	411	19.63
4	Satul	174	8.31
5	Songkla	66	3.15
6	Other	31	1.48
	Total	2,094	100

Table 15. Top five programs in cluster 3

No	Educational institutions	Number	Percent
1	Early Childhood Education	306	14.61
2	Mathematics	171	8.17
3	English and Educational Technology	159	7.59
4	Islamic Studies Teaching	158	7.55
5	General Science	147	7.02

– Cluster 4: Contains 477 students, accounting for 3.60% from various educational institutions, as shown in Table 16.

Table 16. Previous educational institutions for student cluster 4

No	Educational institutions	Number	Percent
1	University in Thailand	180	37.74
2	Secondary Educational Service Area Vocational College	168	35.22
3	Office of the Private Education Commission	52	10.22
4	Department of Learning Encouragement	32	6.71
5	Other	45	9.44
	Total	3,305	100

These educational institutions are in various provinces, as shown in Table 17. Most students have completed the Bachelor's degree, with 188 students accounting for 39.41%, the Science and Mathematics plan, 112 students accounting for 23%, and others with an average grade level of Good (2.5 - 3.4). The top five programs chosen at Yala Rajabhat University as shown in Table 18.

Table 17. Provinces of educational institutions in cluster 4

No	Educational institutions	Number	Percent
1	Bangkok	112	23.48
2	Songkla	82	17.19
3	Mukdahan	61	12.79
4	Nakhon Si Thammarat	24	5.03
5	Pattani	23	4.82
6	Yala	16	3.35
	Other	159	33.34
	Total	477	100

Table 18. Top five programs in cluster 4

No	Educational institutions	Number	Percent
1	Bachelor of Laws	212	44.44
2	Bachelor of Public Administration	34	7.13
3	Teaching Science, Mathematics, and Computer	30	6.29
4	Management	26	5.45
5	Educational Administration	25	5.24

– Cluster 5: Contains 2,009 students, accounting for 15.00% from various educational institutions, as shown in Table 19.

Table 19. Previous educational institutions for student cluster 5

No	Educational institutions	Number	Percent
1	Community College	1,223	60.88
2	University in Thailand	244	12.15
3	Secondary Educational Service Area	172	8.56
4	Office of the Private Education Commission Industrial Community Education College	88	4.38
5	Vocational College	84	4.18
	Other	198	9.85
	Total	2,009	100

These educational institutions are in various provinces, as shown in Table 20. Most students have completed the Vocational Certificate program, with 1,270 students accounting for 38.43%, and the Science and Mathematics plan, 1,021 students, accounting for 30%, with an average grade level of Good (2.5 - 3.4). The top five programs chosen at Yala Rajabhat University as shown in Table 21.

Table 20. Provinces of educational institutions in cluster 5

No	Educational institutions	Number	Percent
1	Yala	691	34.40
2	Narathiwat	474	23.59
3	Pattani	414	20.61
4	Songkla	206	10.25
6	Other	224	11.15
	Total	2,009	100

Table 21. Top five programs in cluster 5

No	Educational institutions	Number	Percent
1	Bachelor of Accounting	584	17.67
2	Business Computing and Digital Technology	463	14.01
3	Management	447	13.53
4	Bachelor of Laws	373	11.29
5	Bachelor of Public Administration	302	9.14

In summary, while DBSCAN provides valuable insights into identifying noise and handling clusters of arbitrary shapes, its higher DBI values in this analysis indicate that it may not be the optimal choice for clustering the given student admission data. Hierarchical clustering performs better but is still outperformed by K-means, which consistently achieves the lowest DBI values. Therefore, K-means is the most suitable clustering algorithm for this application. These clustering outcomes significantly affect Yala Rajabhat University's recruitment and engagement strategies. Understanding the distinct characteristics of each student cluster enables the university to tailor its marketing approaches, scholarship offerings, and academic programs to better meet the specific needs and preferences of different student groups.

3.3. Discussion

The application of clustering techniques has provided valuable insights into student admission patterns [23], [25]. K-means clustering has demonstrated its effectiveness in segmenting students into distinct groups based on admission characteristics, ensuring well-defined clusters [29]. DBSCAN has been particularly useful for identifying outliers and handling variations in density, making it a suitable choice for datasets with irregular cluster distributions [28]. Hierarchical Clustering has further contributed to structuring relationships among student groups, allowing for a hierarchical representation of student segmentation. The DBI has validated that K-means produces the most well-separated and cohesive clusters, reinforcing its reliability in clustering analysis [29]. These findings suggest that clustering techniques can optimize student recruitment strategies by identifying key admission trends and patterns, enabling data-driven decision-making in higher education institutions [23].

A deeper analysis of these findings indicates that clustering techniques offer practical advantages beyond simple segmentation. For instance, K-means clustering enables universities to identify student groups with similar academic performance trends, allowing for personalized academic interventions. DBSCAN, by effectively handling noise and outliers, can reveal hidden trends in student enrollment, potentially identifying at-risk students who do not fit standard clustering patterns. Additionally, Hierarchical Clustering provides an intuitive way to explore relationships among student subgroups, which can be particularly beneficial for institutions seeking to understand long-term enrollment behaviors. By combining these techniques, universities can gain a holistic understanding of student demographics, leading to targeted recruitment and retention strategies.

Beyond recruitment, these clustering methods significantly affect curriculum development, resource allocation, and student support services. Institutions can use clustering insights to design targeted academic programs that align with the strengths and needs of different student groups. Additionally, clustering can help universities allocate financial aid and scholarships more effectively by identifying students who may require additional support based on their academic and socioeconomic backgrounds. Furthermore, integrating clustering with predictive analytics could assist in forecasting student dropouts, enabling early interventions and improved academic guidance.

This study aligns with previous research findings, demonstrating that K-means is highly efficient for structured datasets, while DBSCAN excels in detecting irregular patterns [25]. Unlike prior studies that primarily focused on a single clustering technique, this research presents a comparative evaluation of multiple clustering methods, enhancing its applicability in real-world student admission analysis. The integration of PCA for feature selection has further improved clustering accuracy by reducing dimensionality while retaining key features [28]. PCA also facilitates improved interpretability of clustering results by highlighting the most influential factors in student admission data. However, a key limitation of this study is its reliance on data from a single university, which may impact the generalizability of the findings. Future research should explore larger and more diverse datasets to validate these results and extend the applicability of clustering techniques in higher education data analysis [29]. Additionally, further research could investigate the impact of different feature selection techniques on clustering performance, providing a more nuanced understanding of their effectiveness.

This research contributes to the growing field of educational data mining by demonstrating how clustering techniques can enhance student admissions analysis [23]. The study highlights the importance of data-driven decision-making in university recruitment and academic planning. Despite these contributions, some questions remain unanswered, such as the potential integration of clustering with predictive models to forecast student success. Future studies should explore hybrid models that combine clustering with predictive analytics to offer deeper insights into student behavior and academic outcomes. Additionally, investigating how clustering results evolve over multiple academic years could provide universities with dynamic insights into changing student demographics and enrollment patterns, further enhancing decision-making processes in higher education.

4. CONCLUSION

This study has demonstrated the effectiveness of clustering techniques in analyzing student admission data at Yala Rajabhat University. By employing K-means, DBSCAN, and Hierarchical Clustering, we successfully segmented students into distinct groups based on key admission characteristics. The results indicate that K-means provided the most well-defined clusters, DBSCAN effectively identified outliers, and Hierarchical Clustering allowed for a structured understanding of relationships among student groups. Using the DBI confirmed the clustering validity, reinforcing the importance of selecting the right algorithm for student data segmentation.

The insights gained from this study can significantly contribute to improving recruitment strategies, academic planning, and student support services. Universities can leverage clustering outcomes to develop targeted recruitment campaigns, allocate resources efficiently, and personalize student engagement initiatives. Moreover, applying PCA enhanced clustering accuracy by reducing dimensionality while retaining critical information, ensuring data-driven decision-making in higher education institutions.

Despite its promising findings, this study has some limitations. The analysis was conducted using data from a single university, which may limit the generalizability of the results. Future research should explore the application of clustering techniques across multiple institutions to assess the consistency of the findings. Additionally, integrating predictive models with clustering could provide deeper insights into student success and retention, offering a more comprehensive framework for academic planning and policy-making.

In conclusion, the study underscores the transformative potential of clustering techniques in student admission analysis. By adopting advanced analytics, universities can enhance their strategic decision-making processes, improve student outcomes, and adapt to the evolving educational landscape. Future studies should investigate hybrid models that combine clustering with predictive analytics to offer deeper insights into student behavior and academic outcomes.

APPENDIX

Table 1. Summary of studies on clustering techniques and their results

Study	Methodology/Findings
Li *et al.* [16]	This paper examines the use of clustering technology, particularly the KMEANS algorithm, to analyze student behavior at educational institutions. It highlights how this approach helps segment students into groups for targeted analysis and improved educational management.
Muttaqien *et al.* [17]	Explore the application of the K-means clustering algorithm to enhance the student admission process at Mulawarman University by analyzing data such as GPA, study period, and administrative obedience. The study successfully segments students into different clusters, enabling the university to identify and prioritize new student admissions from feeder schools that historically contributed well-performing students. This methodological approach aims to optimize recruitment strategies and improve the quality of incoming students based on analyzed characteristics. The article's author utilizes data mining techniques, specifically K-means clustering, to analyze factors influencing university admissions. By clustering students based on characteristics like high school GPA and place of residence, the study aims to develop a model to assist in understanding and improving the admission strategies of a university in Southern Thailand. The effectiveness and efficiency of the model are evaluated through clustering data, helping the university align its recruitment activities with the competitive demands of the educational sector.
Kasri and Jati [18]	Combine the K-means clustering algorithm with Simple Additive Weighting (SAW) to optimize university marketing strategies. This combination helps to identify the best locations and strategies for university promotion based on clustering student data, thereby allowing for more targeted and cost-effective marketing approaches.
Rahman *et al.*19]	Focus on developing a course recommendation system for students using K-means clustering and association rule mining. The system utilizes the popularity of courses to generate recommendations, improving course selection for students based on the demand and relevance of the courses. This approach aims to streamline educational pathways and enhance student satisfaction by aligning course offerings with student preferences and career objectives.
Christopher and Edward [20]	Explore using data mining techniques to enhance understanding of student behaviors and academic performance in educational environments. It employs decision trees, association rules, and K-means clustering to evaluate data from newly admitted students, aiming to refine educational strategies and enhance student support through the insights derived from the analysis. A challenge with the K-means clustering algorithm is that it does not inherently specify the optimal number of clusters. To overcome this, the Elbow method can effectively determine the best number of clusters.
Onumanyi *et al.* [21]	Develop an automatic method to determine the optimal number of clusters in a dataset using an improved elbow detection method. This approach helps to automate the selection of cluster numbers for the K-means algorithm, enhancing the accuracy and efficiency of clustering analyses across various data types.
Ghifari and Putri [22]	Developed an innovative method for clustering courses based on student grades using the K-means algorithm, enhanced by the Elbow Method for precise centroid determination. This approach aims to optimize course offerings by identifying which ones consistently yield higher average grades, thereby assisting educational planners and administrators in decision-making processes at educational institutions.
Chapin *et al.* [23]	Develop methods for clustering students' admission data using K-means, hierarchical, and DBSCAN algorithms. Explore how these clustering techniques can help segment students based on their admission data to improve decision-making in academic settings. The study demonstrates the effectiveness of these methods in creating meaningful groupings that can aid universities in optimizing their admissions strategies.
Varna *et al.* [24]	Discuss a framework that combines DBSCAN and K-means clustering algorithms to analyze the behavioral patterns of students from a university in Beijing. The framework utilizes real-time, unbiased behavioral data to efficiently identify anomalous and mainstream student behaviors. This clustering approach provides valuable insights for educational institutions to enhance student management and support services by aligning them more closely with the detected behavioral patterns.
Santosa *et al.* [25]	Use feature selection as part of their methodology to enhance the predictive modeling of students' GPAs. Specifically, they select features based on their relevance to academic performance, such as high school status, location, entrance test scores, and English proficiency. This selection is crucial for effectively clustering students into groups with high and low GPA predictions, enabling more tailored and strategic interventions in the admission process.
Papaioannou *et al.* [26]	Discusses an advanced method for feature selection using parallel computing approaches to enhance the K-means clustering algorithm. This study focuses on reducing the computational time required for feature selection by implementing parallelization techniques. The authors explore different parallelized feature subset selection (FSS) method variations and evaluate their performance on several datasets. The goal is to efficiently manage large datasets by maintaining or improving classification accuracy while significantly reducing execution times. This approach is particularly beneficial when large amounts of data must be processed quickly, as it optimizes data analysis accuracy and speed.
Zhang and Peng [27]	Enhance the traditional K-means approach by embedding a feature selection mechanism to identify the most relevant features for clustering. This integration not only improves the robustness of the clustering process to outliers and noise but also facilitates the discovery of non-linear structures in the data, thereby enhancing the overall effectiveness of the clustering. This method allows for more precise and meaningful clustering by focusing on essential features, thus optimizing the K-means algorithm's performance and interpretability.
Kim *et al.* [28]	Proposed a hybrid Student Dropout Prediction (SDP) model combining XGBoost and CatBoost to improve the precision and recall of university dropout predictions. The model achieved a high precision score of 0.963 and outperformed existing models in recall and F1-score. To analyze the reasons for dropout, the authors applied PCA for dimensionality reduction and K-means clustering, categorizing the causes into four groups: “Employed,” “Did Not Register,” “Personal Issue,” and “Admitted to Other University.” The SDP system enhanced the efficiency of counseling services by accurately identifying high-risk students and providing insights for personalized support.

Table 1. Summary of studies on clustering techniques and their results *(continue…)*

Study	Methodology/Findings
Idrus *et al.* [29]	Explores the effectiveness of different distance measurement techniques in clustering analysis. It focuses on evaluating cluster formation using the K-means and Davies-Bouldin Index (DBI) to determine the optimal cluster configuration. The research examines various distance measures, including Mixed Euclidean Distance, Generalized Divergence, Squared Euclidean Distance, and Mahalanobis Distance, applied to data from villages with primary school facilities in Indonesia. The study underscores the importance of selecting appropriate distance measures for improving clustering accuracy and highlights the practical implications of using clustering techniques for educational and infrastructural analysis.
Cui *et al.* [30]	Uses the HYSPLIT model to generate 72-hour backward air trajectories over Qingdao from 2015 to 2018, then applies three clustering methods—K-means, Hierarchical Clustering, and Self-Organizing Maps (SOM)—to analyze pollutant source pathways. The clustering results are compared using four internal evaluation metrics: DBI, SC, CH, and I index. The findings show that height data contribute little information, so two-dimensional data are sufficient. Among the methods, SOM and K-means perform better than Hier and the default HYSPLIT clustering. The DBI and I index are effective for selecting the optimal number of clusters, with DBI being the preferred choice.

ACKNOWLEDGEMENTS

We thank Yala Rajabhat University's administrative staff for their support and data access, contributing significantly to this research. Our thanks also extend to peer reviewers for their valuable feedback and to our families for their support throughout this project.

FUNDING INFORMATION

The authors state no funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	**C**	**M**	**So**	**Va**	**Fo**	**I**	**R**	**D**	**O**	**E**	**Vi**	**Su**	**P**	**Fu**
Thanakorn Pamutha	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Wanchana Promthong		✓				✓			✓	✓	✓	✓		
Sofwan Palawan	✓		✓	✓			✓	✓		✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

INFORMED CONSENT

This study did not involve human participants or identifiable personal data; therefore, informed consent was not required.

ETHICAL APPROVAL

This study was approved by the Human Research Ethics Committee, Sirindhorn College of Public Health, Yala Province, Thailand, under Approval No. SCPHYLIRB-2567/244.

DATA AVAILABILITY

The data that support the findings of this study are available from Yala Rajabhat University, Thailand. Restrictions apply to the availability of these data, which were used under institutional permission for this study. The data are available from the corresponding author upon reasonable request, provided permission is obtained from Yala Rajabhat University.

REFERENCES

[1] R. Thepprasit and C. Sanrach, "The analysis of factors affecting choosing a major of undergraduate students of the Faculty of Education by using data mining technique," *Journal of Graduate Studies Valaya Alongkron Rajabhat University*, vol. 14, no. 1, 2020.

[2] S. Panpaeng, P. Phanphaeng, J. Kumnuanta, P. Yommakit, K. Kocento, and P. Wongchompoo, "The application of data mining techniques for predicting education to new undergraduate students at Chiang Mai Rajabhat University," in *International Conference on Cybernetics and Innovations, ICCI 2023*, Mar. 2023, pp. 1–6, doi: 10.1109/ICCI57424.2023.10112233.

[3] S. Abadi *et al.*, "Application model of K-means clustering: Insights into promotion strategy of vocational high school," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 2.27 Special Issue 27, pp. 182–187, Aug. 2018, doi: 10.14419/ijet.v7i2.11491.

[4] M. N. Arpay, "Student mining using K-means clustering: A basis for improving higher education marketing strategies," *Psychology and Education: A Multidisciplinary Journal*, vol. 14, pp. 1–8, 2023, doi: 10.5281/zenodo.8383341.

[5] R. M. Cantón-Croda, D. E. Gibaja-Romero, and F. R. Castillo-Villar, "The promotion of graduate programs through clustering prospective students," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 6, pp. 23–32, 2019, doi: 10.9781/ijimai.2019.07.001.

[6] G. A. Sandag, E. Yahuda Putra, R. L. Wurangian, and N. Believer Tulangow, "Analysis of strategy for targeted new student using K-means algorithm," in *2019 1st International Conference on Cybernetics and Intelligent System, ICORIS 2019*, Aug. 2019, pp. 94–99, doi: 10.1109/ICORIS.2019.8874903.

[7] J. Watulangkouw, "Application of data mining to determine promotion strategy using algorithm clustering at SMK Yadika 1," *JISA(Jurnal Informatika dan Sains)*, vol. 5, no. 1, pp. 35–49, Jun. 2022, doi: 10.31326/jisa.v5i1.1107.

[8] A. Heryati and M. I. Herdiansyah, "The application of data mining by using K Means clustering method in determining new students' admission promotion strategy," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 3, pp. 824–833, Feb. 2020, doi: 10.35940/ijeat.c5414.029320.

[9] Z. Zainuddin and A. A. N. Risal, "Balanced clustering for student admission school zoning by parameter tuning of constrained K-means," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, pp. 2301–2313, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp2301-2313.

[10] I. M. Suartana and A. I. N. Hidayat, "Analysis of new student selection using clustering algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 288, no. 1, p. 12079, Jan. 2018, doi: 10.1088/1757-899X/288/1/012079.

[11] W. Prachuabsupakij and S. Chiengpongpan, "Cluster analysis of personal data towards student's graduation in information technology program," in *ACM International Conference Proceeding Series*, Apr. 2020, pp. 76–80, doi: 10.1145/3396743.3396792.

[12] S. Sulastri, L. Usman, and U. D. Syafitri, "K-prototypes algorithm for clustering schools based on the student admission data in IPB university," *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 2, pp. 228–242, Jun. 2021, doi: 10.29244/ijsa.v5i2p228-242.

[13] A. Diana, A. Ariesta, A. Wibowo, and D. A. B. Risaychi, "New student clusterization based on new student admission using data mining method," *Jurnal Pilar Nusa Mandiri*, vol. 19, no. 1, pp. 1–10, Mar. 2023, doi: 10.33480/pilar.v19i1.4089.

[14] R. Andreswari, R. Fauzi, B. M. Izzati, V. P. Widartha, and D. Pramesti, "Students demography clustering based on The ICFL program using K-means algorithm," *International Journal on Informatics Visualization*, vol. 7, no. 2, pp. 555–560, Jun. 2023, doi: 10.30630/joiv.7.2.1916.

[15] Qomariyah and M. U. Siregar, "Comparative study of K-means clustering algorithm and K-Medoids clustering in student data clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 2, pp. 91–99, May 2022, doi: 10.14421/jiska.2022.7.2.91-99.

[16] G. Li, R. Alfred, and X. Wang, "Student behavior analysis and research model based on clustering technology," *Mobile Information Systems*, vol. 2021, pp. 1–6, Nov. 2021, doi: 10.1155/2021/9163517.

[17] H. Muttaqien, M. Lutfi, M. KH, A. Muis, and H. Zainuddin, "Recommendation of student admission priorities using K-means clustering," 2019, doi: 10.4108/eai.2-5-2019.2284614.

[18] M. A. Kasri and H. Jati, "Combination of K-means and simple additive weighting in deciding locations and strategies of University marketing," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 6, no. 2, Oct. 2020, doi: 10.23917/khif.v6i2.11281.

[19] M. M. Rahman, M. S. Islam, R. R. Richi, and A. Chakraborty, "Course course recommendation system for students using K-means and association rule mining," in *ISMSIT 2022 - 6th International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings*, Oct. 2022, pp. 641–646, doi: 10.1109/ISMSIT56059.2022.9932747.

[20] E. Christopher U and A. Edward O, "Clustering, classification, and association rule mining for educational datasets," *International Journal of Advances in Scientific Research and Engineering*, vol. 08, no. 10, pp. 37–51, 2022, doi: 10.31695/ijasre.2022.8.10.4.

[21] A. J. Onumanyi, D. N. Molokomme, S. J. Isaac, and A. M. Abu-Mahfouz, "AutoElbow: An automatic elbow detection method for estimating the number of clusters in a dataset," *Applied Sciences (Switzerland)*, vol. 12, no. 15, p. 7515, Jul. 2022, doi: 10.3390/app12157515.

[22] M. A. Ghifari and W. T. H. Putri, "Clustering courses based on student grades using K-means algorithm with elbow method for centroid determination," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 42–46, Jan. 2023, doi: 10.25139/inform.v8i1.4519.

[23] E. L. Cahapin, B. A. Malabag, C. S. Santiago, J. L. Reyes, G. S. Legaspi, and K. L. Adrales, "Clustering of students admission data using K-means, hierarchical, and DBSCAN algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3647–3656, Dec. 2023, doi: 10.11591/eei.v12i6.4849.

[24] G. Varna, L. Poorvi, and V. S. Priya "An unsupervised ensemble clustering approach for the analysis of student behavioral patterns," *International Journal For Innovative Engineering and Management Research*, pp. 481–490, Sep. 2022, doi: 10.48047/ijiemr/v11/i06/29.

[25] R. G. Santosa, Y. Lukito, and A. R. Chrismanto, "Classification and prediction of students' GPA using K-means clustering algorithm to assist student admission process," *Journal of Information Systems Engineering and Business Intelligence*, vol. 7, no. 1, p. 1, Apr. 2021, doi: 10.20473/jisebi.7.1.1-10.

[26] N. Papaioannou *et al.*, "Parallel feature subset selection wrappers using K-means classifier," *WSEAS Transactions on Information Science and Applications*, vol. 20, pp. 76–86, Mar. 2023, doi: 10.37394/23209.2023.20.10.
[27] Q. Zhang and C. Peng, "Feature selection embedded robust K-means," *IEEE Access*, vol. 8, pp. 166164–166175, 2020, doi: 10.1109/ACCESS.2020.3022749.
[28] S. Kim, E. Choi, Y.-K. Jun, and S. Lee, "Student dropout prediction for University with high precision and recall," *Applied Sciences*, vol. 13, no. 10, Art. no. 6275, 2023, doi: 10.3390/app13106275.
[29] A. Idrus, N. Tarihoran, U. Supriatna, A. Tohir, S. Suwarni, and R. Rahim, "Distance analysis measuring for clustering using K-means and davies bouldin index algorithm," *TEM Journal*, vol. 11, no. 4, pp. 1871–1876, Nov. 2022, doi: 10.18421/TEM114-55.
[30] L. Cui, X. Song, and G. Zhong, "Comparative analysis of three methods for HYSPLIT atmospheric trajectories clustering," *Atmosphere*, vol. 12, no. 6, p. 698, Jun. 2021, doi: 10.3390/atmos12060698.
[31] M. Hamka and N. Ramdhoni, "K-means cluster optimization for potentiality student grouping using elbow method," in *AIP Conference Proceedings*, 2022, vol. 2578, doi: 10.1063/5.0108926.
[32] S. Liu, S. Cao, M. Suarez, E. C. Goonetillek, and X. Huang, "Multi-Level DBSCAN: A hierarchical density-based clustering method for analyzing molecular dynamics simulation trajectories," Jun. 2021, doi: 10.1101/2021.06.09.447666.

BIOGRAPHIES OF AUTHORS

Thanakorn Pamutha received his M.Sc. degree in Computer Science from Prince of Songkla University, Songkhla, Thailand, and his Ph.D. degree in Information Technology from the Faculty of Information Technology, Rangsit University, Thailand. He has interests in database systems, artificial intelligence, data mining, and machine learning. He formerly held the position of Director of the Office of the President at Yala Rajabhat University (YRU), Thailand, from 2013 to 2023. He was also the Assistant Rector at YRU from February 2023 to May 2023. He is currently an Assistant Professor in the Department of Information Technology, Faculty of Science, Technology, and Agriculture at Yala Rajabhat University (YRU), Thailand. He can be contacted via email at thanakorn.p@yru.ac.th.

Wanchana Promthong received his M.Sc. degree in Information Technology Management, specializing in Computer and Network Systems, from the Faculty of Engineering at Prince of Songkla University, Thailand, in 2014. He also holds a B.Sc. degree in Business Computer from the Faculty of Commerce and Management at Prince of Songkla University, Trang Campus, which he obtained in 2005. His research interests include network security systems and the application of information technology. He is currently working as a Computer Technical Officer at the Office of Academic Resources and Information Technology, Yala Rajabhat University (YRU), Thailand. He can be contacted via email at wanchana.p@yru.ac.th.

Sofwan Pahlawan received the B.Sc. degree in Information Technology from the Faculty of Information Technology, Rangsit University, Thailand. He has interests in database systems, Web Application Development. He formerly held the position of computer scientist at the Office of the President at Yala Rajabhat University (YRU), Thailand. He is currently a computer scientist in the Department of Information Technology for Administration, Office of Academic Resource and Information Technology Center at Yala Rajabhat University (YRU), Thailand. He can be contacted via email at sofwan.p@yru.ac.th.