DOI: https://doi.org/10.55164/ecbajournal.v16i3.268195

Unveiling Student Segments: Leveraging Clustering Analysis of Registration Data for Enhanced Recruitment Strategies

Prajak Chertchom[1*]

Received : 23/06/2023
Revised : 25/07/2023
Accepted : 20/08/2023

Abstract

Recruiting the right students is crucial for higher education institutions to achieve their goals of attracting talented individuals and fostering academic success. This study aims to investigate the effectiveness of data-driven recruitment strategies in identifying target student populations and tailoring recruitment efforts accordingly. By utilizing K-means clustering analysis on a dataset of student profiles, this research identifies distinct clusters based on factors such as province of residence, parent occupation, and parent revenue. The findings reveal valuable insights into the characteristics and preferences of different student clusters, enabling the development of targeted recruiting strategies. The results indicate that a significant number of students in specific clusters are from Songkhla Province, have parents predominantly engaged in the private sector, agriculture, and fishery occupations, and come from families with moderate income levels. Building upon these findings, several recommendations for recruiting strategies are proposed. These include focusing marketing efforts in the identified regions, forging partnerships with local businesses, offering financial aid programs, and establishing connections with agricultural and fishing communities. This can involve targeted online advertisements, promotional campaigns in local schools, and active participation in career fairs or educational expos held in Songkhla Province. This study contributes to the field of higher education recruitment by leveraging data-driven approaches to identify target student populations and develop tailored strategies.

Keywords: K-mean, Clustering Analysis, Registration Data, Recruitment Strategies, Customized Approaches

Introduction

In today's competitive educational landscape, institutions face the challenge of attracting and enrolling the right students who align with their mission, programs, and values. The traditional one-size-fits-all recruitment approach is no longer sufficient in effectively engaging with diverse student populations. Educational institutions need to adopt targeted strategies that resonate with specific student segments to maximize their recruitment efforts. The abundance of registration data provides a valuable resource for

[1] Assistant Professor Dr., Faculty of Economics and Business Administration, Thaksin University
* Corresponding author: E-mail address: prajak.c@tsu.ac.th

understanding students' preferences, behaviors, and interests. By leveraging the power of clustering analysis, educational institutions can uncover distinct student segments within their registration data, allowing them to tailor their recruitment strategies and communication approaches accordingly. (Chiu & Chai, 2020; Muhajir et al., 2022) The current state of higher education in Thailand underscores the urgency of adapting universities to accommodate a larger student population. In each TCAS (Thai University Central Admission System) cycle from 2018 to 2022, the number of students admitted falls short of the universities' actual capacity. With a total of 155 universities, encompassing both public and private institutions, Thailand has the potential to accommodate approximately 420,000 new students each year. However, statistics from the TGAT/TPAT64 (Thai General Aptitude Test/Thai Professional and Academic Test) exam reveal that there are only around 257,282 applicants, indicating a significant shortfall in the number of prospective students. This calls for the implementation of measures to ensure universities can effectively cater to the increasing demand for higher education. (MHESI, 2023) Clustering analysis is a data exploration technique that groups similar data points into clusters based on their characteristics or patterns. Applied to student registration data, clustering analysis allows for the identification of student segments with similar registration patterns, course preferences, or academic performance metrics. This analysis helps institutions gain insights into the diverse needs and motivations of students, enabling them to develop targeted recruitment strategies that resonate with each segment. By employing clustering analysis, institutions can move beyond generic marketing campaigns and instead create personalized recruitment approaches. For example, they can develop specific messaging for high-achieving students who tend to enroll in advanced courses or design targeted outreach programs for students interested in specific fields of study. (Davies & Bouldin, 1979; Asif & Haider, 2017) This customized approach not only enhances the institution's appeal to potential students but also improves the efficiency of recruitment efforts by focusing resources on the most promising segments. Understanding and leveraging student segments through clustering analysis can also aid in the optimization of resource allocation. By identifying clusters of students with shared characteristics, institutions can allocate resources such as recruitment personnel, financial aid, or specialized programs more effectively. (Bandyapadhyay et al., 2023).

Objective

1. To apply the K-Mean clustering technique to students' registration data for the purpose of identifying distinct student segments.
2. To explore the characteristics and patterns of each identified student segment based on registration data.

3. To examine the potential implications of the identified student segments on recruitment strategies in educational institutions.

Literature review

The researcher has conducted an extensive study, reviewing textbooks and relevant documents covering the following topics:

1. Data Mining

Data Mining involves the exploration of vast datasets to uncover concealed associations and forecast trends. It delves into the realm of Knowledge Discovery in Databases (KDD), aiming to extract novel insights from extensive data collections. This process tackles the analysis of substantial data volumes to unearth hidden patterns and relationships within the dataset. Additionally, it incorporates the integration of diverse scientific fields, such as databases and technologies like Online Analytical Processing (OLAP), which facilitate the organization of extensive business databases. This technology encompasses the storage, collection, and preparation of data for utilization in data mining. It involves the application of machine learning algorithms to uncover hidden patterns and relationships within the data. Statistical and data analysis methods are employed for initial data exploration, revealing concealed patterns and relationships. This process incorporates various disciplines such as information science, mathematical programming, and high-performance computing. Due to its time-consuming nature, efficient calculations are essential to support data mining. Visualization techniques are employed to present the results, formats, and relationships of the data in a user-friendly manner. This aids in interpreting and applying the outcomes to business contexts. The Cross-Industry Standard Process for Data Mining (CRISP-DM) is widely recognized as the most popular data mining process. It serves as a framework for translating business problems into distinct stages of the Data Mining process, independent of technology application and usage. This framework comprises six steps, as illustrated in Figure 1. (Ayele, 2021; Plotnikova et al., 2022)

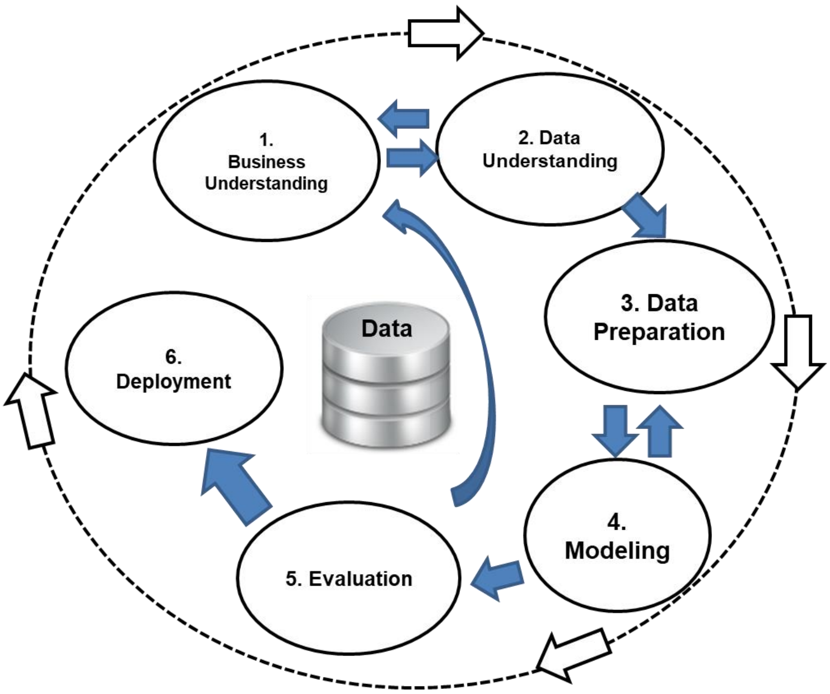


Figure 1 Cross-Industry Standard Process for Data Mining Model (The CRISP-DM model)

Therefore, in today's customer-centric era of business, it is crucial to possess comprehensive information that enables accurate tracking, analysis, and prediction of customer needs. This empowers businesses to drive continuous growth, even amidst external fluctuations and multifaceted risk factors. By applying the insights gained from this information, organizations can optimize their strategies and maximize their own benefits.

2. Overview of Clustering Analysis

The K-means Clustering technique is employed to calculate the distances between data points and group them based on their highest similarity within the same cluster. It involves dividing the data into K groups, with K representing the desired number of clusters. The process begins by selecting initial centroids for each cluster. Next, the distances between each data point and the centroid of each cluster are calculated, and the data is assigned to the cluster with the highest similarity. The algorithm then computes new centroids for each cluster by taking the average of the grouped data within each cluster. This iterative process continues, refining the grouping of data points. K-means Clustering finds applications in data clustering across various fields (Frades & Matthiesen, 2010; King, 2015), including the classification of disease risk groups. Despite its popularity and simplicity, k-means has a limitation: it is suitable for numerical feature segmentation. One of the strengths of K-means Clustering is its high computational speed, making it well-suited for analyzing large volumes of business data. It can quickly group information and provide easily interpretable results. The outcomes of k-means clustering offer straightforward data clusters that users can readily interpret.

The flexibility and customizability of k-means Clustering are additional advantages. The technique can be customized by adjusting various parameters to enhance clustering efficiency, such as selecting better starting points. However, k-means Clustering has its weaknesses. The number of clusters (K) needs to be predetermined by the analyst, and specifying the desired number of clusters is essential before running the algorithm. If the results do not align with the specified number of clusters, readjustments are necessary. Moreover, k-means clustering may yield inconsistent or different results with each adjustment. (Goyal & Vohra, 2012; Fahad et al., 2014; Asif et al., 2017)

The K-means Clustering technique offers numerous applications in business. For instance, businesses can employ it to cluster customers based on their buying behavior and characteristics. By considering various data points like age, gender, location, income, and preferred product types, businesses gain valuable insights into the behavior and needs of different customer groups. This enables them to tailor their strategies and offerings accordingly. Another application is Bank Account Opening Data Analysis, where banks can analyze customer data related to account openings. By examining factors such as age, income, account types, and other banking services utilized, banks can understand customer behavior and enhance their services to meet customer needs effectively. Grouping this data facilitates a comprehensive analysis of customer behavior patterns, enabling banks to make informed decisions. Furthermore, K-means Clustering is widely used for product grouping and analysis. By applying this technique, businesses can group similar products together, enabling them to gain a deeper understanding of customer buying patterns and trends. This knowledge helps in refining product offerings, designing targeted marketing strategies, and improving overall customer satisfaction. (Papamitsiou & Economides, 2014). By leveraging K-means Clustering in these areas, businesses can gain valuable insights into customer behavior, improve service offerings, and enhance their overall understanding of market dynamics.

3. Previous Research on Clustering Analysis

The research paper titled "Educational Data Mining for Prediction of Student Performance Using Clustering Algorithms" explored the utilization of clustering algorithms in educational data mining to forecast student performance. The main objective of the study was to uncover patterns and trends within student data by employing clustering algorithms and utilize this knowledge to make predictions about academic achievements. The authors specifically focused on investigating the efficacy of two clustering algorithms, namely K-means and Expectation-Maximization (EM), in predicting student outcomes. (Durairaj & Vijitha, 2014) To conduct the research, a dataset comprising student attributes such as age, gender, family background, and previous academic records were utilized. The clustering algorithms were then applied to group students based on these attributes, enabling the identification of distinct clusters and patterns within the dataset. The study concluded that both K-means and EM algorithms are effective in predicting student performance. The clustering approach offered valuable insights into the factors that influence academic outcomes. By analyzing the identified clusters, the researchers were able to make predictions about student performance and identify potential areas for intervention or support. The research developed a trustworthy model using data mining techniques that can extract essential information, thereby suggesting the adoption of this model as a strategic management tool within the present education system.

The second paper titled "Educational Data Mining: A Survey and a Data Mining-based Analysis of Recent Works" provided a comprehensive survey of educational data mining (EDM) and analyzes recent works in the field. The study explored the state of the art in EDM and identify the trends, techniques, and applications that have emerged in recent years. The author examined a wide range of research articles and publications related to EDM and conducts a systematic analysis of their methodologies and findings. In addition, the paper covered various aspects of EDM, including data collection and preprocessing, feature selection and extraction, classification and prediction, clustering, association rule mining, and visualization techniques. Furthermore, the author discussed the different approaches used in these areas and highlights their strengths and limitations. (Peña-Ayala, 2014) In summary, the paper provided insights into the application of EDM in different educational domains, such as e-learning, intelligent tutoring systems, educational assessment, and student performance prediction. The author examines the methodologies and results of several studies in these domains to showcase the practical applications of EDM. The third paper titled focused on the application of clustering algorithms in educational data mining. (Dutt et al, 2015) The study explored the use of clustering algorithms to analyze educational datasets and extract meaningful patterns and insights. The researchers investigate various clustering algorithms, including K-means, DBSCAN, and Hierarchical Clustering, to identify their effectiveness in educational data mining. By applying these clustering algorithms to educational datasets, the researchers grouped students based on their similarities in attributes such as academic performance, demographic information, and learning styles. This grouping allows for the identification of clusters or subgroups of students with similar characteristics or behaviors.

The research findings demonstrated the effectiveness of clustering algorithms in educational data mining. Clustering algorithms help identify student profiles and patterns that can inform decision-making processes in education. These insights can be utilized to personalize instruction, identify at-risk students, and improve teaching and learning strategies. The forth paper titled "A Systematic Review on Educational Data Mining" provided a comprehensive overview of the field of educational data mining through a systematic review. (Dutt et al., 2017) The study examined the existing literature on educational data mining and provide insights into the various techniques, methodologies, and applications in this field. The authors conduct an extensive review of research articles, conference papers, and books published between 2007 and 2016, focusing on key themes and trends in educational data mining. The systematic review identifies several prominent areas of research within educational data mining, including student performance prediction, student profiling, personalized learning, and educational recommender systems. The authors analyzed the methodologies and algorithms employed in these studies and highlight the strengths and limitations of each approach. The review also discusses the challenges and future directions of educational data mining, such as privacy concerns, data quality issues, and the need for interpretability and transparency in the analysis process. The research highlighted the immense potential of educational data mining in providing valuable insights for educational institutions and stakeholders to enhance teaching and learning processes.

The fifth paper titled "Analysis of Educational Data Mining" explored the field of educational data mining and its applications. (Ahuja et al., 2019) The study focused on analyzing educational data using data mining techniques to gain insights and make informed decisions in the field of education. The authors discussed various data mining methods and algorithms that are commonly employed in educational data mining, including decision trees, clustering, association rule mining, and classification. The research also highlighted the importance of educational data mining in improving the teaching and learning processes. It emphasized the potential of data mining techniques to uncover patterns, trends, and relationships within educational datasets, which can help educators and stakeholders make data-driven decisions and enhance educational outcomes. In addition, the paper provided an overview of educational data mining techniques, discussed their strengths and limitations, and presented real-world examples of their applications. Finally, the authors emphasized the need for further research and development in this field to fully harness the potential of educational data mining for educational improvement.

A summary of the papers on educational data mining, along with a table presenting the key information is as following;

Table 1 Family Income Transformation

Paper	Focus	Techniques	Findings
Durairaj, M., & Vijitha, C. (2014)	Clustering algorithms for predicting student performance	K-means, EM clustering	K-means and EM clustering were effective in prediction
Dutt, A., Aghabozrgi, S., Ismail, M. A. B., & Mahroeian, H. (2015)	Clustering algorithms applied in educational data mining	K-means, Hierarchical, EM clustering	Various clustering algorithms applied in educational mining
Dutt, A., Ismail, M. A., & Herawan, T. (2017)	Potential of educational data mining for improving teaching and learning	N/A	Educational data mining provides valuable insights
Jeyaraj A., & Kumar S.S. (2018)	Data clustering for predicting student performance	K-means, Hierarchical, DBSCAN	K-means clustering outperformed other techniques
Peña-Ayala, A. (2014)	Educational data mining: A survey and a data mining-based analysis of recent works	N/A	Provides a comprehensive survey and analysis of recent works in educational data mining

Methodology

The scope of this quantitative research encompasses the application of data mining techniques, specifically K-Means Clustering, to extract knowledge from large datasets. The research focuses on utilizing these techniques in the context of a bachelor's degree program.

1. Scope of research

1.1 Scope of Content: this research investigates the factors influencing students' decisions to pursue higher education at the university level. By employing data mining tools and techniques, the study analyzes data collected from the student registration system. The focus is primarily on exploring relationships and patterns among various variables, including high school grade point averages, province of residence, and family status.

1.2 Scope of population and data collection

The study utilizes student data from the registration system of a government university in the southern region. The population for this research comprises data from the academic year 2019 to the academic year 2021. A total of 12,125 data were selected for analysis in this study.

1.3 Data analysis tools and techniques

The researcher adopted the clustering technique as a data mining approach, specifically utilizing the K-Means Algorithm. This technique involves grouping data into smaller clusters based on their similarities, without any prior knowledge of the underlying groups within the data. Clustering is widely used in data analysis, particularly when dealing with large datasets. It enables the researcher to analyze and organize data with similar characteristics into distinct groups, facilitating further analysis and interpretation of the data. The K-Means Algorithm is employed to iteratively assign data points to clusters based on the minimization of distances between data points and cluster centroids. $d\,(x,y) = \sqrt[2]{\sum_{i=1}^{n}(x_i\ ,\ \ y_i)^2}$ (Celebi, 2014) Furthermore, the researcher employs four essential research tools for data analysis: Excel, Power BI, WEKA, and Rapid Miner. By leveraging these four research tools, the researcher can effectively handle various aspects of data analysis, including data manipulation, visualization, exploration, and advanced modeling techniques. In the data analysis process, data preparation plays a crucial role in ensuring the suitability of data for analysis. It involves several key steps, as outlined by Ayele (2020) These processes encompass data collection, data inspection, data cleaning, data selection, and data transformation. In this specific research, the researcher converts the data into numerical features to facilitate analysis using the K-means Clustering technique, as demonstrated in Table 2 and Table 3. By diligently following these data preparation steps, the researcher ensures the data's accuracy, completeness, and suitability for subsequent analysis. This rigorous approach helps in maintaining data integrity and enables meaningful insights to be extracted from the data during analysis.

Table 2 Family Income Transformation

Range Of Income	Family Income Indeed
<= 15,000 baht per month	1
15,001 - 30,000 baht per month	2
30,001 - 45,000 baht per month	3
45,001 - 60,000 baht per month	4
no income	5

Table 3 Sample of Province Transformation

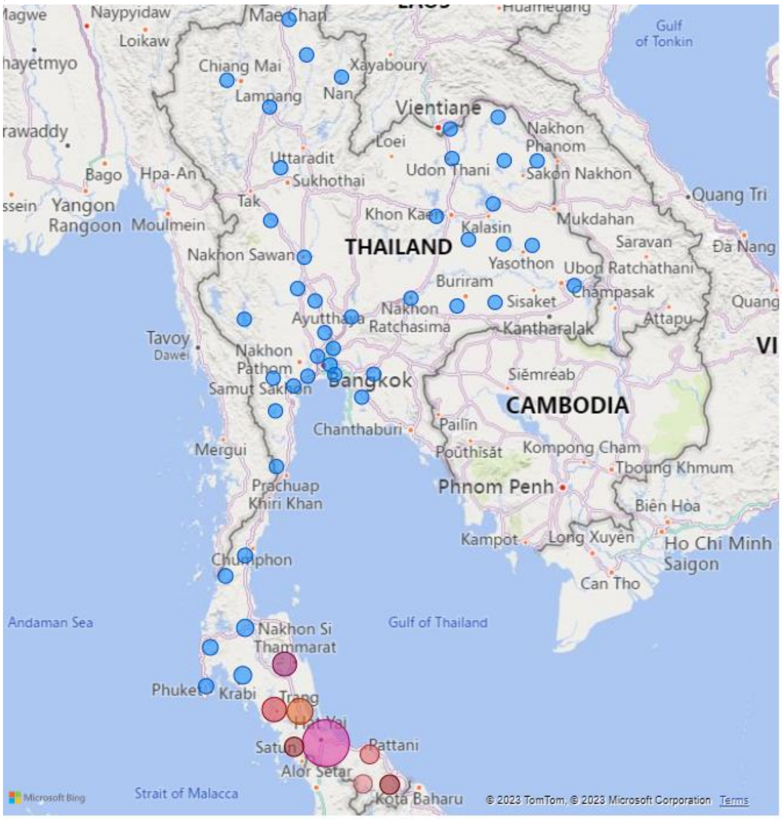
Province	Province Index No.
Bangkok	1
Buriram	2
Chiang Rai	3
Chonburi	4
Chumphon	5
Phuket	17
Songkhla	25
Surat Thani	26
Surin	27
Trang	28
Udon Thani	29
Yala	30

Based on Table 2, the researcher performed two data conversions. Firstly, the provinces where the students resided were translated into English. This conversion was to facilitate the use of software for generating visualizations, making the process more convenient and efficient. By converting the province names into English, the researcher ensured compatibility with the software and streamlined the data processing stage. This approach allowed for smoother visualization creation and enhanced data processing capabilities.

Results

Descriptive Analytics

The following figure represents the number of students by province from the years 2019 to 2021. Each row in the table under figure represents a specific province or category, and each column represents the corresponding number of students for each year.



Songkhla	4197
Phatthalung	1474
Trang	1298
Nakhon Si Thammarat	1251
Narathiwat	681
Satun	654
Pattani	599
Yala	540
Krabi	442
Surat Thani	377
Phangnga	169
Phuket	130
Chumphon	108
Ranong	66
Bangkok	24
Chiang Rai	12
Nonthaburi	9
Chiang Mai	6
Phetchaburi	6
Ratchaburi	6
Khon Kaen	5
Nakhon Ratchasima	5
Phayao	5
Prachuap Khiri Khan	5
Udon Thani	5
Chonburi	4
Pathum Thani	4
Roi Et	4
Ubon Ratchathani	3
Buriram	2
Maha Sarakham	2
Sakon Nakhon	2
Samut Prakan	2
Total	**12123**

Figure 2 Number of Students by Province from 2019 -2021

Figure 2 illustrates the enrollment statistics for the academic years 2019-2021, depicting a total of 12,123 students. These students are distributed across various provinces of residence in Thailand. The data, compiled using POWER BI, offers valuable insights into the population of students in each province. The graph depicts a regional map of Thailand, where each province is represented by a symbol in the form of a circle.

The size of each circle corresponds to the number of students residing in that province. Larger circles indicate a higher number of students. From the graph, it is apparent that the province with the highest number of students during the academic years 2019-2021 is Songkhla, followed by other provinces in Thailand. It highlights the five provinces with the highest student enrollment, along with their respective student counts. In addition, it is evident from the table that Songkhla has the highest number of students, with 4,197 students in residence during the academic years 2019-2021. Phatthalung follows closely behind with 1,474 students. Trang, Nakhon Si Thammarat, Narathiwat, Pattani, and Yala provinces also have significant student populations, with 1,298, 1,251, 681, 599, and 540 students, respectively.

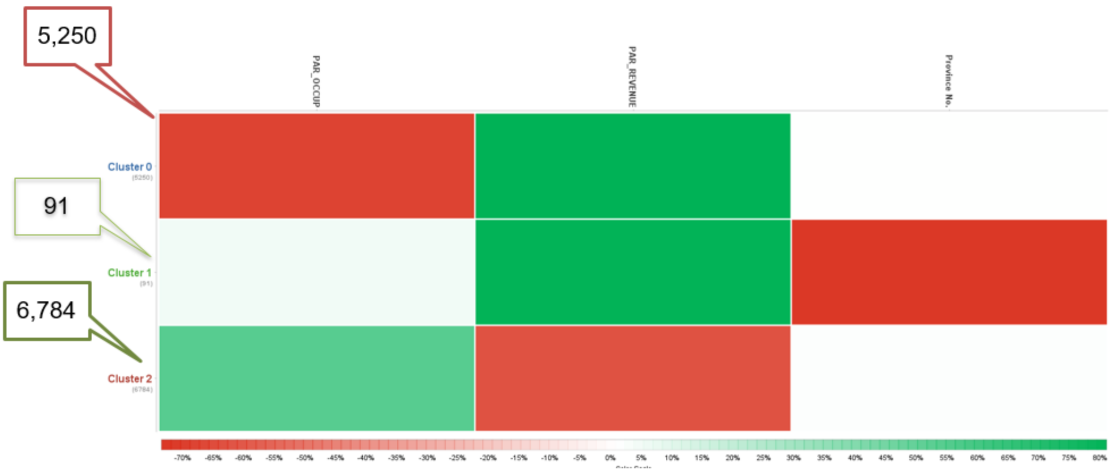


Figure 3 a Heat Map of K-means Clustering

From figure 3, a heat map of k-means clustering represents the visual representation of the clustering results using colors to indicate the similarity or dissimilarity between data points. Figure 2 presents the specific values associated with each cluster, which are 5,250 students for Cluster 0, 91 for Cluster 1, and 6,784 students for Cluster 2.

Cluster	PAR_OCCUP	PAR_REVENUE	Province No.
Cluster 0	1.988	3.890	6.151
Cluster 1	4.352	3.879	1.626
Cluster 2	5.919	1.579	6.194

Figure 4 K-means Centroid

Figure 4 provides essential insights into the centroids of three distinct clusters: Cluster 0, Cluster 1, and Cluster 2. Each cluster is characterized by specific attribute values. In terms of "Parent Occupation," Cluster 0 has a value of 1.988, Cluster 1 has a value of 4.352, and Cluster 2 has a value of 5.919. Similarly, for the "Parent Revenue" attribute, Cluster 0 has a value of 3.890, Cluster 1 has a value of 3.879, and Cluster 2 has a value of 1.579. Lastly, the "Province" attribute highlights values of 6.151 for Cluster 0, 1.626 for Cluster 1, and 6.194 for Cluster 2. These centroid values play a crucial role in understanding the unique characteristics and features that distinguish each cluster within the dataset.

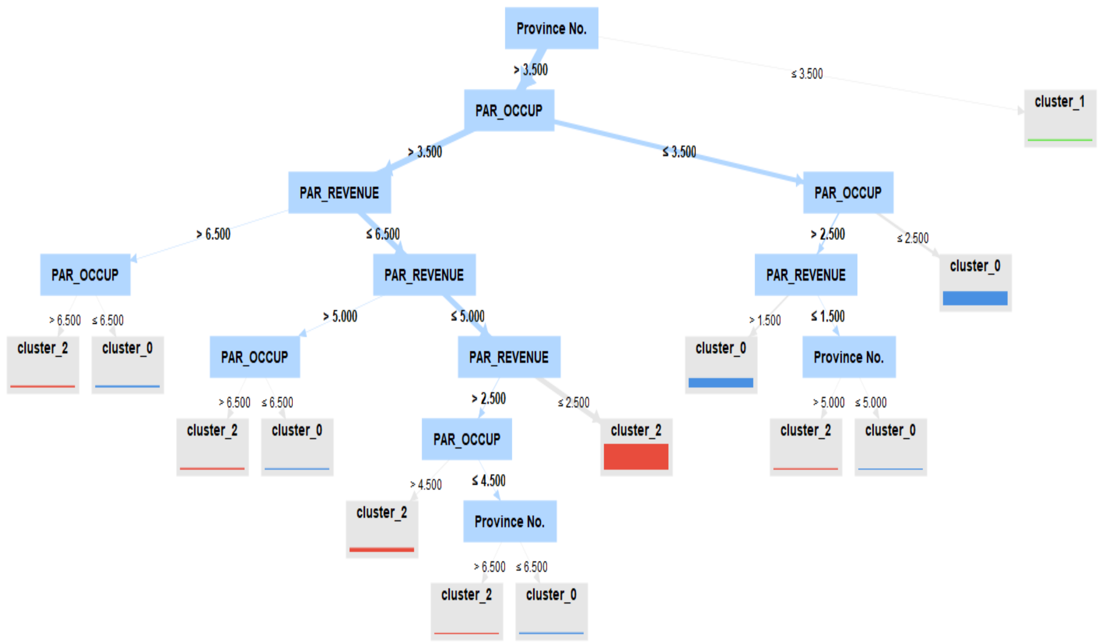


Figure 5 K-means Cluster Tree

Figure 5 illustrates a hierarchical tree-like structure that organizes clusters and their sub-clusters, providing valuable insights into their relationships. The structure begins with the parent node "Province," followed by "Parent Occupation," and culminates in the root node "Parent Revenue." Initially, the dataset is divided into clusters based on the provinces associated with the data points. Within each province cluster, the K-means algorithm proceeds to subdivide the data based on the occupation of the parents. Lastly, within each sub-cluster of parent occupations, further clustering takes place, this time based on the parent revenue. This hierarchical approach enables a comprehensive exploration of the data, facilitating the understanding of how clusters and sub-clusters relate to one another based on the attributes of "Province," "Parent Occupation," and "Parent Revenue." The tree suggests that students' university choices are influenced first by "Province," followed by "Parent Occupation," with the ultimate decision resting on the root node "Parent Revenue."

Discussion

Based on Figure 2, it can be inferred that the researcher divided students based on factors influencing their decision to study at the observed university, as part of objective number 1 of this research. The clustering experiment resulted in three distinct groups. Among the students who chose to study in the academic years 2019-2021, cluster 2 stood out with the highest number of data points, consisting of 6,784 students primarily from Songkhla Province. The average value for the "Province" attribute in this group was 6.194, indicating that a majority of the students hailed from Songkhla Province. (The transformed number 6 represents Songkhla in the data preparation process.) Furthermore, the "Parent Occupation" attribute had an average value of 5.919, suggesting that most of the students' parents worked in the private sector. (The transformed number 5 represents parents employed in the private sector.) Additionally, the average value for the "Parent Revenue" attribute was 1.579, indicating that the majority of students' families had a monthly income ranging from 15,001 to 30,000 baht. (The transformed number 1 represents an average revenue of 15,001 to 30,000 baht per month.) Therefore, based on these findings, it can be concluded that the university's primary customer base comprises students residing in Songkhla Province. These students come from families predominantly employed in the private sector, with a monthly income range of 15,001 to 30,000 baht.

The second group, Cluster 0, comprised the second-largest number of data points, totaling 5,250 students. Within this group, the average value for the "Province" attribute was 6.151, indicating that a significant number of students were from Songkhla Province. In terms of the "Parent Occupation" attribute, the average value was 1.988, suggesting that a majority of the students' parents were engaged in agriculture and fishery occupations. Furthermore, the average value for the "Parent Revenue" attribute was 3.890, signifying that most of the students' families had a monthly income ranging from 45,001 to 60,000 baht. Therefore, within Cluster 0 and Cluster 2, it is evident that a considerable proportion of students hailed from Songkhla Province. Their parents' occupations predominantly included employment in the private sector, as well as the agriculture and fishery sectors. Moreover, these students belonged to families with a moderate monthly income range of 15,000 to 60,000 baht.

Suggestion

Suggestions for applying the research results

Based on the findings and observations, the following recruitment strategies can be implemented for the university. Firstly, considering the substantial number of students in Cluster 0 and Cluster 2 hailing from Songkhla Province, it is advantageous to concentrate marketing efforts in this region. This can involve targeted online advertisements, promotional campaigns in local schools, and active participation in career fairs or educational expos held in Songkhla Province. By increasing visibility and engagement in the area, the university can attract more prospective students.

Secondly, given that Cluster 0 exhibits a high proportion of students with parents employed in the private sector, establishing partnerships with local businesses and industries can greatly enhance recruitment endeavors. Collaborative initiatives such as internships, guest lectures, or industry-specific programs that align with the interests of potential students can provide them with valuable experiences and create a mutually beneficial relationship between the university and the private sector.

Thirdly, recognizing that a significant portion of students in Cluster 0 and Cluster 2 come from families with moderate income levels, offering scholarships and financial aid programs can improve access to education and attract talented students who require financial support. Emphasizing the university's commitment to affordability and the availability of financial assistance options can be a compelling message during recruitment efforts.

Furthermore, considering the notable presence of parents engaged in agriculture and fishery occupations within Cluster 0, establishing connections with local farming and fishing communities can be advantageous. This can involve designing specialized courses or research opportunities in areas such as agricultural sciences, aquaculture, or environmental studies to attract students interested in these fields. By aligning the university's offerings with the needs and aspirations of prospective students from these backgrounds, recruitment efforts can yield positive results. By implementing these strategies, the university can effectively target and engage prospective students, building a diverse and talented student body while establishing valuable partnerships with various sectors of the community.

The findings of this study align with the research conducted by Goyal and Vohra (2012) in the field of higher education, which investigated the diverse applications of data mining techniques. The results from our study support the notion that data mining can be effectively employed to extract valuable insights and patterns from extensive educational datasets. This, in turn, contributes to enhancing student retention, conducting comprehensive educational data analysis, and ultimately improving decision-making processes for better educational outcomes. (Dutt & Herawan, 2017)

Recommendations for future research or further research

Based on the findings of this study, several potential directions for future research can be pursued. Firstly, it is recommended to investigate the long-term impact of the recruitment strategies implemented based on the cluster analysis. This could involve assessing the retention rates, academic performance, and career outcomes of students recruited through these targeted strategies, comparing them to those recruited through traditional methods. Understanding the effectiveness and sustainability of these strategies in the long run is crucial for improving recruitment practices.

Secondly, further exploration into the factors influencing student decisions to choose a specific university or program is warranted. This research could delve deeper into various aspects such as academic reputation, geographical location, curriculum offerings, extracurricular activities, and facilities provided by the university. By gaining a comprehensive understanding of these factors, institutions can tailor their recruitment efforts to align with the preferences and needs of prospective students.

Lastly, it would be valuable to conduct comparative research by comparing the findings of this study with similar studies conducted in other universities or institutions. This comparative analysis can help identify common trends, best practices, or unique patterns that may provide additional insights into effective recruitment strategies. By examining the experiences and outcomes in different contexts, a more comprehensive understanding of successful recruitment approaches can be achieved.

By addressing these future research areas, institutions can refine their recruitment strategies, enhance student engagement, and ultimately improve their overall recruitment outcomes.

References

Ahuja, R., Jha, A., Maurya, R., & Srivastava, R. (2019). *Analysis of Educational Data Mining. In Harmony Search and Nature Inspired Optimization Algorithms: Theory and Applications, ICHSA 2018*. Springer Singapore.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing Undergraduate Students' Performance using Educational Data Mining. *Computers and Education*, 113, 177-194.

Ayele, W. Y. (2020). Adapting CRISP-DM for Idea Mining: a Data Mining Process for Generating Ideas using a Textual Dataset. *International Journal of Advanced Computer Sciences and Applications*, 11(6), 20-32.

Bandyapadhyay, S., Fomin, F. V., Golovach, P. A., Lochet, W., Purohit, N., & Simonov, K. (2023). *How to Find a Good Explanation for Clustering? Artificial Intelligence*, Volume 322, 103948. https://doi.org/10.1016/j.artint.2023.103948

Celebi, M. E. (Ed.). (2014). Partitional Clustering Algorithms. *Springer*, 147-192

Chiu, T. K. F., & Chai, C. S. (2020). Sustainable Curriculum Planning for Artificial Intelligence Education: A Self-Determination Theory Perspective. *Sustainability*, 12 (14), Article 5568. https://doi.org/10.3390/su12145568

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE transactions on Pattern Analysis and Machine Intelligence*, (2), 224-227.

Dutt, A., Aghabozrgi, S., Ismail, M. A. B., & Mahroeian, H. (2015). Clustering Algorithms Applied in Educational Data Mining. *International Journal of Information and Electronics Engineering*, 5(2), 112.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*, 5, 15991-16005.

Peña-Ayala, A. (2014). Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works. *Expert Systems with Applications*, 41(4), 1432-1462.

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., & Bouras, A. (2014). A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Transactions on Emerging Topics in Computing*, 2(3), 267-279.

Goyal, M., & Vohra, R. (2012). Applications of Data Mining in Higher Education. *International Journal of Computer Science Issues (IJCSI)*, 9(2), 113.

King, R. S. (2015). *Cluster Analysis and Data Mining: An introduction*. Mercury Learning and Information.

Muhajir, D., Akbar, M., Bagaskara, A., & Vinarti, R. (2022). Improving Classification Algorithm on Education Dataset using Hyperparameter Tuning. *Procedia Computer Science*, 197, 538-544.

Ministry of Higher Education, Science, Research and Innovation (MHESI). (2023). *Number of New students, Academic Year 2021, Semester 1 Classified by Province*. https://info.mhesi.go.th/stat_std_new.php?search_year=2564&download=7085&file_id=202208151109.xlsx.

Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A systematic Literature Review of Empirical Evidence. *Journal of Educational Technology & Society,* 17(4), 49-64.

Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM Data Mining Process in the Financial Services Industry: Elicitation of Adaptation Requirements. *Data & Knowledge Engineering*, 139, 102013. https://doi.org/10.1016/j.datak.2022.102013