**ISSN** 1816-6075 (Print), 1818-0523 (Online)
**Journal of System and Management Sciences**
Vol. 12 (2022) No. 3, pp. 156-180
DOI:10.33168/JSMS.2022.0309

# Geo-marketing Promotional Target Selection using Modified RFM with Spatial and Temporal Analysis: A Case Study

Ernawati[1], Safiza Suhana Kamal Baharin[2], Fauziah Kasmin[2], Heribertus Andre Purwanugraha[1]

[1] Universitas Atma Jaya Yogyakarta, Indonesia
[2] Universiti Teknikal Malaysia Melaka, Malaysia

ernawati@uajy.ac.id (corresponding author); safiza@utem.edu.my; fauziah@utem.edu.my; andre.purwanugraha@uajy.ac.id

**Abstract.** This study proposes a framework for assisting decision-makers in selecting a target market using a Recency-Frequency-Monetary (RFM) based model. This study aims to identify a university's promotional target market for student recruitment and develop a geo-marketing strategy. This study employed a modified RFM model tailored to the university's marketing context, combining it with spatial analysis to quantify feeder schools "value" and temporal analysis to examine the enrollment patterns of the highest value feeder schools using data mining techniques and Geographic Information System (GIS). A case study of an Indonesian university was used in this study. This study identified 108 (4.63%) out of 2,334 feeder schools as the prioritized target market, contributing 51.54% of 18,537 enrolled students to the university during the analysis period. The prioritized feeder schools are in 32 cities and 23 regencies from 368 cities/regencies, with the majority being private schools. The research's findings revealed the distribution of feeder schools in regencies/cities and the trend of enrolled students from the highest value feeder schools segment, which can assist university management in selecting target feeder schools more precisely. Based on the findings, decision-makers can create a geo-marketing strategy for promotional activities and direct resources to the prioritized feeder schools. This study contributes by reinforcing a modified RFM model with spatial and temporal analysis to help university decision-makers choose feeder schools as the university's target market and develop a geo-marketing strategy.

**Keywords:** data-mining, geo-marketing, RFM-model, spatial-temporal-analysis, target-selection, university-promotion

# 1. Introduction

Higher education institutions (HEIs) face increasing competition to attract students. To overcome the competition, institutions should apply marketing strategies (Pardiyono et al., 2022). Developing an effective marketing strategy in a competitive environment is critical (Sukoroto et al., 2020), as is promotion, one of the marketing elements. Promotion is an essential element in marketing that aims to inform, persuade, or remind the target audience about the products/services (Lamb et al., 2011). Besides the quality of the information and the communication channels used, who receives the information is critical in promotional activities. Hence choosing the right target market is one of the important marketing decisions (Aghdaie & Alimardani, 2015) for HEIs to increase prospective students. A well-targeted marketing campaign can boost recruitment success while saving resources.

Senior high schools are university feeder schools since most universities recruit their graduates. Finding appropriate feeder schools for a target market helps HEIs recruit students. The "value" of a feeder school to a university must be quantified when determining the target market. The Recency-Frequency-Monetary (RFM) model is widely used to express customer value (Abbasimehr & Shabani, 2019; Chiang, 2017) and determine a target market (Hosseini & Mohammadzadeh, 2016; Hwang & Lee, 2021). The original RFM model employs the recency variable to represent the newness of customer relationships with the business, the frequency variable to measure the number of transactions made by customers, and the monetary variable to count the amount of money paid by customers in a given period. RFM analysis can help businesses better understand the profitability of their customers (Beheshtian-Ardakani et al., 2018). The RFM model is advantageous for the target customer selection model because it captures customer characteristics with few features, reducing model complexity and improving model transparency (Chen et al., 2012).

Although the RFM model measures customer values based on customer profitability to the business, the RFM model has been modified for application in educational institutions, which are non-profit institutions. The RFM model was redefined to segment high schools based on their loyalty to the university (Hidayat et al., 2020). The study defined recency as the number of times a specific school's alumni registered in the university during the analysis period, frequency as average registrants from the school every year, and monetary variable as the total registrant from the school during the analysis period. The research used the Fuzzy C-means algorithm to cluster high schools with partnerships with the university based on the new student admission data and found four potential partner categories. Moreover, the RFM model was modified to the Recency-Length-Frequency-CGPA-Persistence (RLFCP) model for quantifying feeder school value as a target market (Ernawati et al., 2022a). Recency and length variables identify the relationship value. Recency (R) was defined as the novelty of a feeder school's last alumnus enrolled at the university; Length (L) measures the length of time since the first and last alumni of a feeder school enrolled at the university in the analysis period; Frequency (F) indicates the number of feeder school's alumni enrolled at the university. The cumulative grade point average (CGPA) and Persistence (P) show the quality of a feeder school's alumni when studying at the university until the end of the fourth semester. The K-means algorithm and Elbow method grouped the feeder schools into clusters. The feeder school clusters were ranked based on their weighted RLFCP, and the weight variable was calculated using the Full Consistency Method (FUCOM). As a result, four clusters with the highest value were set as the target segment. However, the studies did not include location information data, which is important in marketing (Libório et al., 2020).

Data is an essential source of information in developing a marketing strategy and can be used in modern management to support optimal decision-making (Mekvabidze, 2020). By empowering data mining techniques, businesses can discover more valuable insights into their customers, which further helps understand customers better (Dahiya et al., 2021). About 80% of business-relevant data is related to spatial issues (Cliquet & Baray, 2020). Spatial data contains information relating to the location of an object on the earth's surface. The use of spatial data in marketing raises a popular term called geo-marketing. Geo-marketing refers to the involvement of spatial data with the company's internal and external data using the Geographic Information System (GIS) as a tool (Verschuren, 2006) to expand and strengthen marketing planning and implementation activities (Suhaibah et al., 2016) and assist in management decision making (Melnyk L. & Nyzhnyk L., 2018). Geo-marketing has a big impact on business. It improves a firm's marketing strategies by analyzing customers' characteristics in a particular location and developing a plan that optimizes promotion and reaches customers (Guarda & Augusto, 2019). Combining a marketing database with geo-marketing can help marketers quickly visualize and define key client segments and develop an effective marketing strategy (Vitor et al., 2018).

There are various geo-marketing applications, such as location and target market analysis. Location analysis is used to identify the most suitable location, such as school locations (Constantinidis, 2019), pharmacies (Chacón-garcía, 2017), and supermarkets (Baviera-Puig et al., 2016), while geo-marketing target analysis identifies where the high concentration of potential customers is located and then defines the target groups. Nunes et al. (2014) sought a location for a new outdoor advertisement for a food company. They determined which area has a high concentration of people who fit the target market before deciding on a location. The study used socioeconomic data from a 25-49-year-old housewives survey to identify areas with the highest potential consumers and visualized it using a GIS tool. According to the findings, the target audience and sales increased due to the new outdoor advertising installation. Surjandari & Rosyidah (2017) used GIS to conduct spatial analysis to determine the distribution of fixed broadband customers. The study employed Kernel Density Estimation to display a heatmap of customers. The research discovered seven customer density centers scattered throughout the study area, with the highest density at the main center 3. However, these previous studies did not use the RFM model.

Ernawati et al. (2022b) extended the RFM model with spatial analysis to discover target segments using university enrollment and school spatial data. The RFM model was modified by adapting its variables to the marketing context of the educational institution. The modified RFM model was enhanced with a district's potential (D) variable resulting from the spatial analysis and K-means clustering. The proposed model (RFM-D) and the K-prototype algorithm determined the target segment. The study set the two highest-value clusters of five as the target segment. Compared to the CLV-based RFM, the proposed model's performance was better. This research only extended the RFM model with a spatial analysis model, while customer characteristics change over time. According to Morris & Thrall (2010), enrollment is a trend, so longitudinal data should be analyzed to detect changes in the enrollment pattern.

Several previous studies have used RFM to determine the dynamic behavior of customers. Abbasimehr & Shabani (2021) predicted a bank's customer behavior at the segment level by using one variable of RFM, namely the monetary variable. The study proposed a combination of time series clustering and forecasting methods. Time-series clustering was used to investigate the time-series customer behavior feature. Following that, time-series forecasting was used to forecast the future behavior of customer segments. The study's results showed that the combined method outperforms all other individual forecasters. Hosseini & Shabani (2015) classified customers by incorporating time and trend of customer value change over several periods. By splitting time into separate periods, each customer value for each time interval based on the RFM model was calculated. The K-means algorithm was applied to classify customers in each time interval, yielding a changing value trend for each customer. Again, K-means was used to cluster the changing value trends of customers. After applying it to business-to-business (B2B) purchasing transaction data, the research discovered 11 clusters with similar value-changing trends. However, these studies only focused on temporal data, not combining it with spatial data. Kelly (2019) suggested considering the temporal dimensions of spatial data. Ernawati et al. (2021) conducted a data-driven spatial-temporal analysis to determine potential university promotion activity regions. The study employed two-stage clustering and GIS-based multi-criteria decision-making (MCDM) to discover the potential regions. As a result, the regencies and cities that consistently contributed many high-quality students were chosen as priority regions. However, this study did not apply the RFM model. Therefore, this study will improve the prior studies. This research aims to propose a framework that integrates spatial and temporal analysis to the modified

RFM model for selecting a promotional target market and use it to recommend a geo-marketing strategy based on a case study of an Indonesian university.

The contribution of this study is proposing a new framework for supporting decision-makers in determining a target market for university promotional activities. This study proposes a modified RFM model integrated with spatial analysis for determining a feeder school value and incorporating temporal analysis for analyzing the enrollment pattern of the highest value feeder school cluster. This study took a case study on an Indonesian university. According to the case study results, a recommended geo-marketing strategy is suggested. This paper is organized as follows: Section 2 describes the material used and the proposed framework. Section 3 shows the case study results, followed by a discussion. Finally, a brief conclusion is presented in Section 4.

# 2. Material and Method

## 2.1. Data and study area

This study used enrollment and academic data from an Indonesian university as a case study. The university is in a regency of Daerah Istimewa Yogyakarta Province on Java Island. Since many students from outside the province attended the university, the study area of this research covers Indonesia country. The spatial unit for analysis was the district administrative (in Indonesia, called regency/city). The Indonesian districts and provinces' shapefiles were acquired from the Indonesian Geospatial Portal. Indonesia has 514 districts spread across 34 provinces. The provinces are distributed on five main islands, namely: Sumatra Island (eight provinces with 140 districts), Kalimantan Island (five provinces, 56 districts), Sulawesi Island (six provinces, 81 districts), Java Island (six provinces, 119 districts), Papua Island (two provinces, 42 districts), and four archipelagos, namely: Riau Archipelago (one province with seven districts), Bangka Belitung Archipelago (one province with seven districts), Nusa Tenggara Archipelago (three provinces, 41 districts), Maluku Archipelago (two provinces, 21 districts) (BPS-Statistics Indonesia, 2019).

## 2.2. The proposed framework

The proposed framework is illustrated in Fig. 1. There are five phases in the framework as follows:

(1) Phase 1: Data acquisition.

The first phase begins with acquiring internal data from the university database and external data from Google Maps and related geospatial portals.

(2) Phase 2: Data preprocessing.

Data preprocessing aims to prepare data to be ready to be processed. The preprocessing performed includes data integration, cleaning, selection, transformation, and aggregation. The collected data from various sources was integrated and cleaned to detect errors, overcome incomplete data, and remove redundant data. The relevant features were selected, transformed, and aggregated. The aggregation was performed on the enrolled students based on their high school and the high schools by their regency/city.

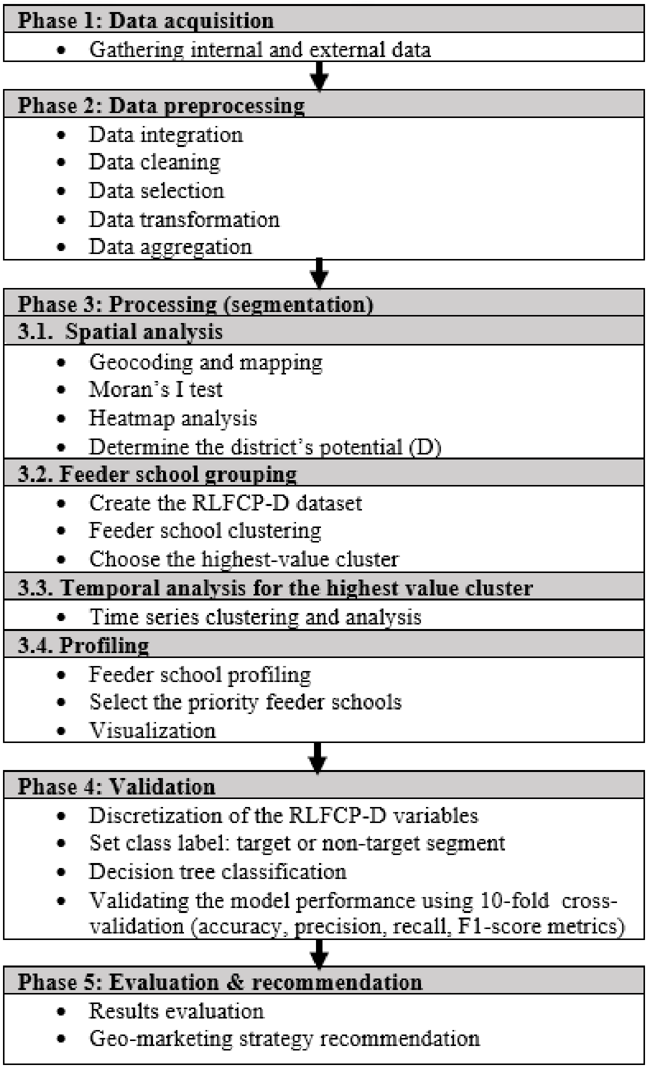



Fig. 1: The proposed framework

(3) Phase 3: Segmentation.

This phase deals with feeder school segmentation. As shown in Fig. 2, there are four major steps for feeder school segmentation: a spatial analysis based on the school's regency/city, feeder school grouping based on the RLFCP-D model, a temporal analysis of the most valuable feeder schools' cluster, and profiling to determine the priority feeder-schools for the target market. The steps are discussed in detail below.

**Spatial analysis.** Spatial analysis was conducted to determine the distribution of feeder schools by regency/city and the number of enrolled students from each high school and district. Feeder school locations were geocoded first, and then the number of enrolled students was mapped at the origin school level. After that, the Global Moran's I spatial autocorrelation test was employed (Ghodousi et al., 2020) to test the randomness distribution of the feeder schools in the districts. If the observed value (I) of the Moran index exceeds the expected value (E(I)), the hypothesis that there is random spatial distribution is rejected, implying there were spatially clusters. Subsequently, the heatmap analysis was performed (Surjandari & Rosyidah, 2017) to visualize the location's concentration of the feeder schools. The spatial analysis was also used to determine the potential of a feeder school district as a promotional target. The feeder school district's potential (D) as a university promotional target was categorized based on the number of feeder schools and the enrolled students at the university from the district using the K-means method (Ernawati, 2022b). Since the K-means method requires the optimal number of clusters as input, the Elbow method combined with the Silhouette index was used to determine the optimal number of clusters (k) (Kit et al., 2021). Before applying K-means, the data was normalized using Z-score normalization.

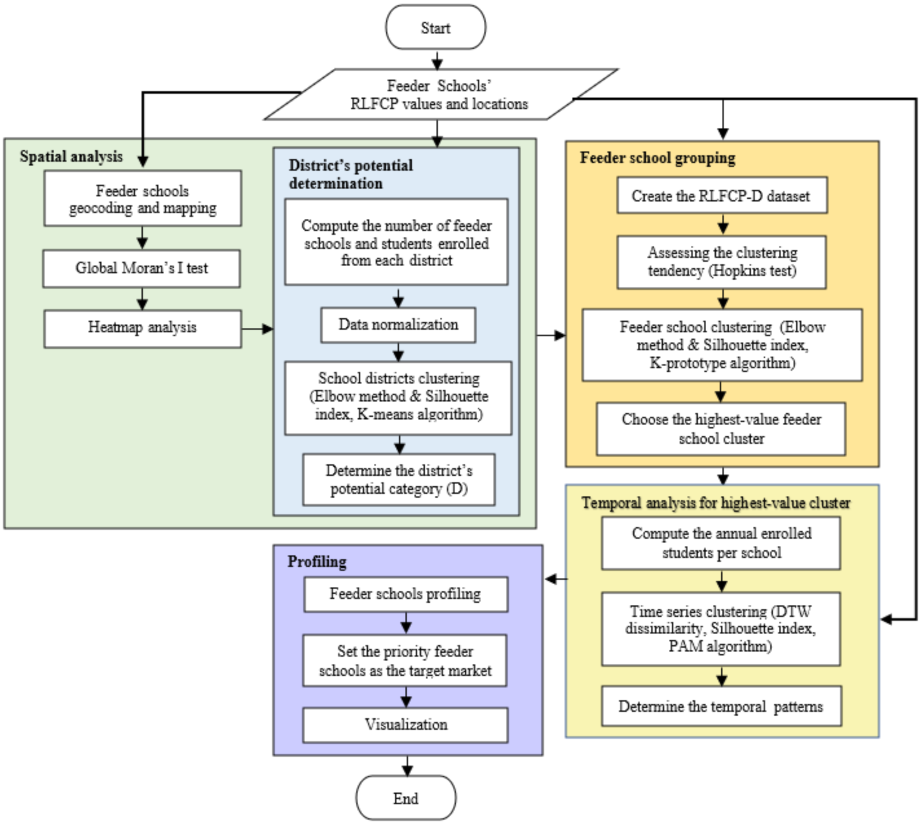



Fig. 2: The flowchart of the segmentation phase

**Feeder school grouping**. This study integrated the RLFCP model (Ernawati et al., 2022a) with the D variable obtained from the previous spatial analysis step (Ernawati, 2022b). Table 1 presents the description of the RLFCP-D variables used in this study. For the feeder school grouping step, the RLFCP-D dataset was first prepared. Then the Hopkins test was applied to assess the clustering tendency after normalizing the RLFCP-D variables using Z-score normalization (Lawson & Jurs, 1990). Following that, the RLFCP normalized values and the D value were clustered using K-prototype (Huang, 1998; Sulastri et al., 2021), a mixed-type data clustering algorithm, because the district's potential (D) is a categorical variable, while others are numerical. For determining k, a combination of the Elbow method and the Silhouette index was used. Then, the highest-value cluster was selected as the target segment.

Table 1: The RLFCP-D model description

| Variable | Description |
|---|---|
| Recency (R) | The novelty of a feeder school's most recent alumnus enrolled at the university during the analysis period, on a scale from 1 to 9. |
| Length (L) | Time since the first and last feeder school graduates enrolled at the university. On a scale of 0 to 8. |
| Frequency (F) | The total number of the feeder school alumni enrolled at the university during the analysis period. |
| CGPA (C) | The average CGPA of a feeder school's alumni at the university's end of the fourth semester. On a scale of 0 to 4. |
| Persistence (P) | The percentage of the feeder school alumni who persist until the end of the fourth semester, on a scale of 0 to 100. |
| District's potential (D) | The potential category of a feeder school district, on a scale of 1 to 4. |

**Temporal analysis for the highest value cluster.** This study conducted a temporal analysis to identify the enrollment patterns of the highest-value cluster of feeder schools. Each school's annual enrolled students numbers were calculated, and temporal analysis was performed using time series clustering. The Partitioning Around Medoids (PAM) algorithm with Dynamic Time Warping (DTW) dissimilarity measure was adopted to reveal the enrollment pattern. The PAM clustering method is one of the k-medoids algorithms commonly used for time series data clustering (Li et al., 2020), whereas DTW is the most popular and accurate shape-based dissimilarity method (Aghabozorgi et al., 2015; Javed et al., 2020). Since PAM belongs to partitional clustering, same as K-means and K-prototype, k was required to run the algorithm. The Silhouette index was used to determine k values, and then the highest-value cluster members were divided into k different enrollment patterns (Abbasimehr & Shabani, 2021).

**Profiling.** Based on the previous steps' results, the feeder school cluster profiles were identified, and the prioritized feeder schools as the target market were defined. Using GIS tools, the locations of priority feeder schools were visualized, thus helping decision-makers to determine geo-marketing strategies for promotional activities.

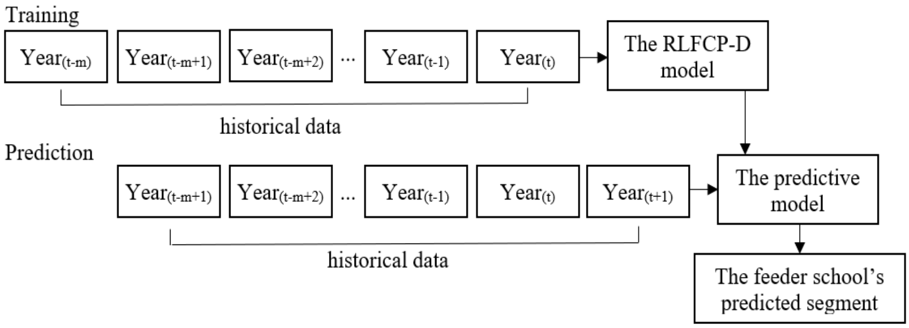
(4) Phase 4: Validation.

The fourth phase was performed to validate the proposed framework using the classification technique (Hosseini & Mohammadzadeh, 2016). The Classification and Regression Tree (CART) algorithm was used to construct a predictive decision tree model based on the clustering results after the numerical variables were discretized into two categories: above the variable's average value (A) and the same or lower than the average value (B). As for the tree's leaves, two-class labels: target segment (TS) and non-target segment (NT), were used. Then, 10-fold cross-validation (Bunnak et al., 2015) was conducted to compute accuracy, precision, recall, and F1-score metrics (Cheng & Morimoto, 2019; Ernawati et al., 2022a).

(5) Phase 5: Evaluation and recommendation.

Since the RLFCP-D model is time-related, adopting the model from Levin & Zahavi (2005), the predictive model was then used for predicting the next academic year's feeder schools' segment, as shown in Fig. 3. This study compared the predictive model's output with the real class label obtained from the university's promotion office to evaluate the RLFCP-D model performance. Finally, a recommendation based on the research findings was proposed to support university management in determining the target market and developing a geomarketing strategy.

Fig. 3: Feeder school's segment prediction process




# 3. Results and Discussion

## 3.1. Data acquisition and preprocessing

For a case study, this study used enrollment and academic data for the 2010-2018 batch from a university located in Daerah Istimewa Yogyakarta Province, on Java Island, Indonesia. There were 21,476 enrolled student records. For spatial analysis and visualization in the GIS environment, the Indonesia province and regency shapefile were obtained from the Indonesian Geospatial Portal, while high schools' latitude and longitude were collected from Google Maps.

This study only examined students from high schools in Indonesia. Hence students from out of Indonesia high schools and records that did not have information of high school origin were ignored. The preprocessing conducted includes cleaning, selection, integration, aggregation, and transformation. Data cleaning was performed to clean dirty data, namely detecting errors in the data, overcoming incomplete data, and removing redundant data. All of the dirty data were excluded from the data target for analysis. After cleaning, this study included 18,537 students' records for analysis and then selected variables used to derive RLFPC variables and spatial data. Data integration on student data with origin high school spatial data was carried out. Data aggregation was performed on the enrolled students based on their high school, yielding 2,334 feeder schools. Following that, the feeder schools were aggregated by their regency/city. While data normalization used before the clustering process was Z-score normalization.

## 3.2. Spatial analysis

As a result of geocoding to map the feeder schools' location on the Indonesia district and province map layer, Fig. 4 shows their distribution and contribution to the university's enrollment. Each green bubble represents a feeder school, and its size reflects the number of students enrolled proportionally. The map shows that the feeder schools are spread throughout Indonesia. The feeder schools were distributed in 368 (71.60%) of Indonesia's districts, spanning all provinces. They were located in 276 of 416 regencies and 92 of 98 cities. Out of 18,537 students, 39% were from feeder schools in regencies, and 61% were from feeder schools in cities.


Fig. 4: The feeder schools' distribution

Although distributed in many regions, the results of the spatial autocorrelation test on the number of feeder schools in the districts using Moran's Global I test showed that the Moran's index I = 0.3663, E(I) = -0.0028 with pseudo-p-value = 0.001 (less than 0.05). This study concluded that the feeder schools are clustered because I > E(I), implying the null hypothesis that feeder schools are randomly distributed in the regencies/cities is rejected. Fig. 5 visualizes the classification of the district based on the number of feeder schools. From the five main islands, most of the districts in Java and Kalimantan had feeder schools, but many districts in Sumatra, Sulawesi, and Papua did not, as indicated by the grey area. Most districts had few feeder schools, between one to 12 feeder schools. Of the top ten districts with the highest number of feeder schools, seven were on Java Island, where the university is located.

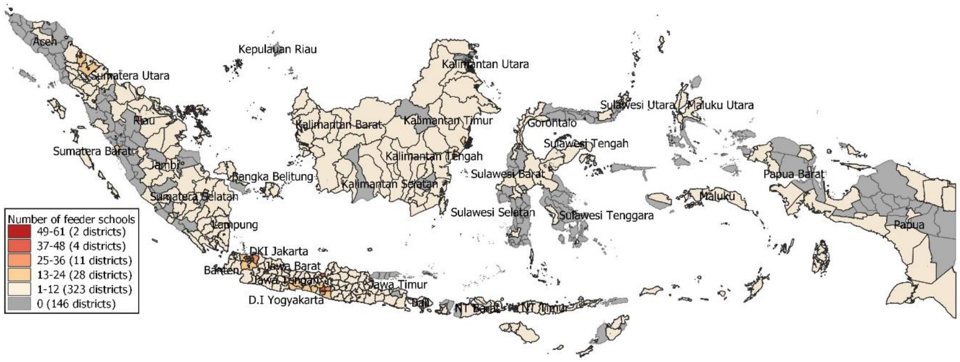



Fig. 5: District classification based on the number of feeder schools

Fig. 6 presents a heatmap of the number of feeder schools in districts. It shows most feeder schools were concentrated on Java Island, in the districts surrounding the Daerah Istimewa Yogyakarta Province, where the university is located. Besides that, the feeder schools were also more densely in the districts in Central Java (Jawa Tengah) Province, the nearest province to the Daerah Istimewa Yogyakarta Province and around the Capital City of Indonesia (DKI Jakarta).


Fig. 6: Feeder schools’ concentration

The result of the heatmap analysis is supported by Fig. 7, which depicts the district classification based on the number of feeder schools on Java Island. As shown by the color of the districts, most of the districts with many feeder schools were located on Java Island.


Fig. 7: Java's districts classification based on the number of feeder schools

Fig. 8 depicts the classification of districts based on the number of enrolled students at the university from each district. Most districts sent fewer than 50 students during the nine academic years. Of the top ten districts, three were in Daerah Istimewa Yogyakarta Province, and five were in Jawa Tengah Province. Both of the provinces are on Java Island.

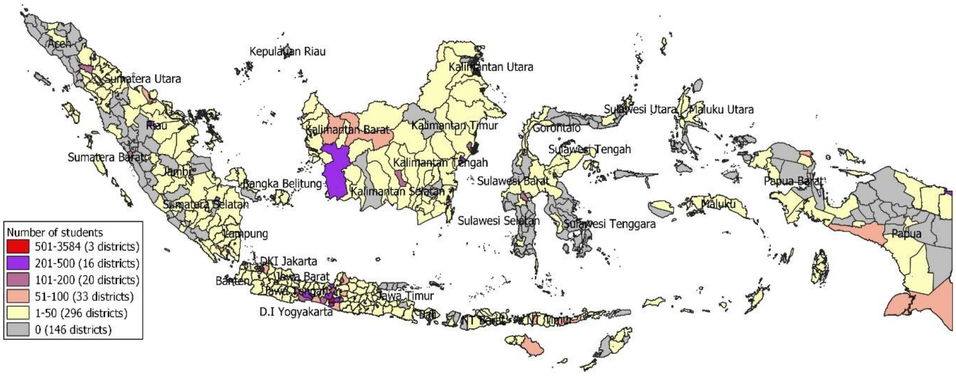



Fig. 8: District classification based on the number of enrolled students

Fig. 9 highlights Java districts. The top three Indonesian districts as the student sources of the university were on Java Island, while six districts still did not send enrolled students to the university.

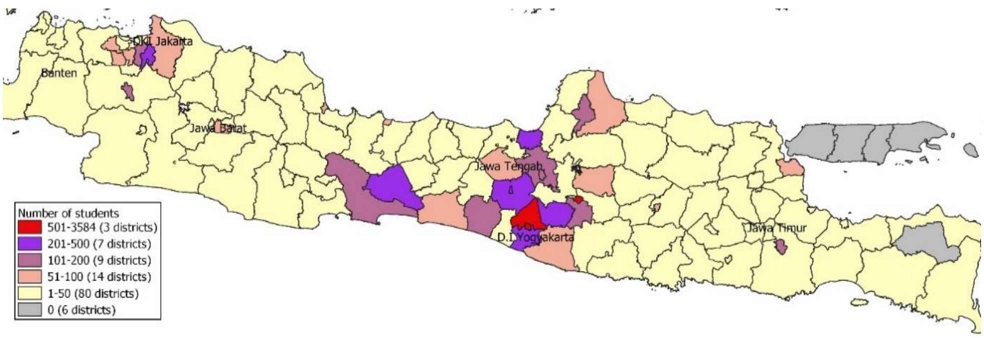



Fig. 9: Java's districts classification based the number of enrolled students

Out of 18,537 enrolled students at the university in the 2010-2018 period, most were from feeder schools in the city (61.21%). Meanwhile, based on the school category, most of the regency school's students were from public schools (53.67%), while the city school's students were from private schools (69.97%). The thematic map of the number of enrolled students from each feeder school is displayed in Fig. 10. According to the map, the contribution of each feeder school varies. Most feeder schools are one-time feeders (indicated by a green bubble). Therefore, each feeder school's value needs to quantify to determine which feeder schools are valuable for the university as the target market, and the potential of the school district needs to know as information in decision making.

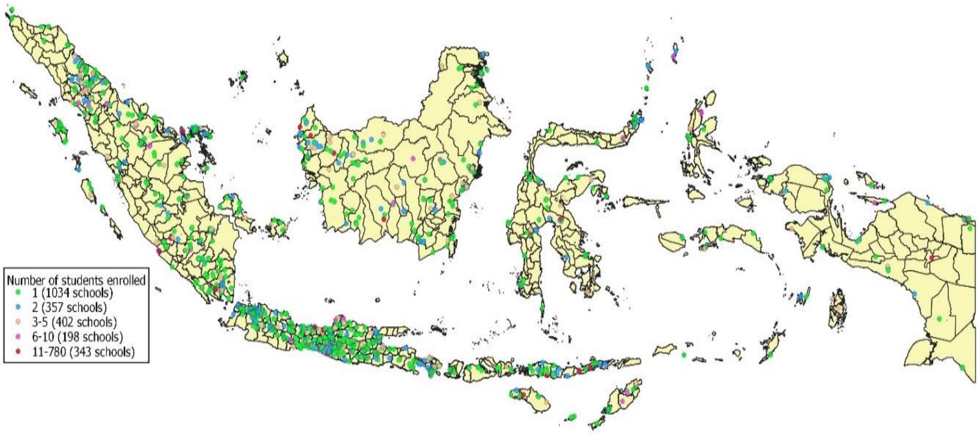



Fig. 10: School classification based on the number of enrolled students

Since the number of feeder schools and the number of enrolled students from each district differed, districts' potential was not the same. Thus, this study conducted clustering to determine the potential of a district as a university target market. Table 2 shows the four categories of district potential after performing K-means clustering. Before that, data normalization was performed, and the optimal number of clusters (k) was determined using the Elbow method and the Silhouette test. K=4 was chosen with Silhouette index = 0.6999. In the table, the cluster's center X shows the average number of enrolled students, and the cluster's center Y shows the average number of feeder schools from each district.

Table 2: District’s potential clusters

| Cluster | Cluster’s center | | Number of Districts | Number of students | Number of schools |
|---|---|---|---|---|---|
| | X | Y | | | |
| D1 | 13.86 | 3.39 | 300 | 4157 | 1016 |
| D2 | 111.60 | 15.07 | 55 | 6138 | 829 |
| D3 | 388.17 | 35.67 | 12 | 4658 | 428 |
| D4 | 3584 | 61 | 1 | 3584 | 61 |

The four categories of the potential district are the lowest potential (D1), moderate (D2), good (D3), and the best (D4). The higher the district's potential, the more feeder schools and students enrolled at the university from the district. There are 300 districts in D1, with 1,016 feeder schools that sent 4,157 students. On average, each district has three feeder schools and contributes 14 students. There are 55 districts included in D2, with 829 feeder schools supplying 6,138 students. D3 has 12 districts that sent 4,658 students from 428 feeder schools. At the same time, D4 contains one city with 61 feeder schools contributing 3,584 students. The city is in Daerah Istimewa Yogyakarta Province, nearest to the university's regency.

The spatial analysis results show that more cities are included in higher potential areas than regencies, and most city school students are from private schools. In Indonesia, cities' educational facilities and education levels are generally better than in regencies. Thus, it is reasonable if the proportion of cities included in the higher potential districts is greater. There are, on average, four feeder schools in each regency and twelve feeder schools in each city, with six enrolled students from regency feeder schools and ten students from city feeder schools.

## 3.3. Feeder schools grouping

Table 3 shows an example of RLFCP-D data used for feeder school grouping. The dataset contains 2,334 feeder schools for training.

Table 3: The RLFCP-D dataset

| No | Feeder_School_Code | R | L | F | C | P | D |
|---|---|---|---|---|---|---|---|
| 1 | 11.04.AN.001 | 3 | 2 | 2 | 2.33 | 100 | 1 |
| … | … | … | … | … | … | … | … |
| 1354 | 34.71.AN.006 | 9 | 8 | 65 | 3.05 | 98.46 | 4 |
| … | … | … | … | … | … | … | … |
| 2334 | 92.71.MN.010 | 6 | 0 | 1 | 1.92 | 100 | 2 |

The result of the Hopkins statistic was 0.9594, greater than 0.75. It was concluded that the dataset has a high clustering tendency. Thus, the feeder schools were divided into five clusters using the K-prototype algorithm with a Silhouette value of 0.4068. Fig. 11 visualizes the clusters' features resulting from K-prototype clustering, while Table 4 presents the characteristics of each cluster, which are shown by the average value of each normalized-RLFCP variable, distribution of feeder schools in each category of the district potential (D), number of feeder schools, and the number of the enrolled students.

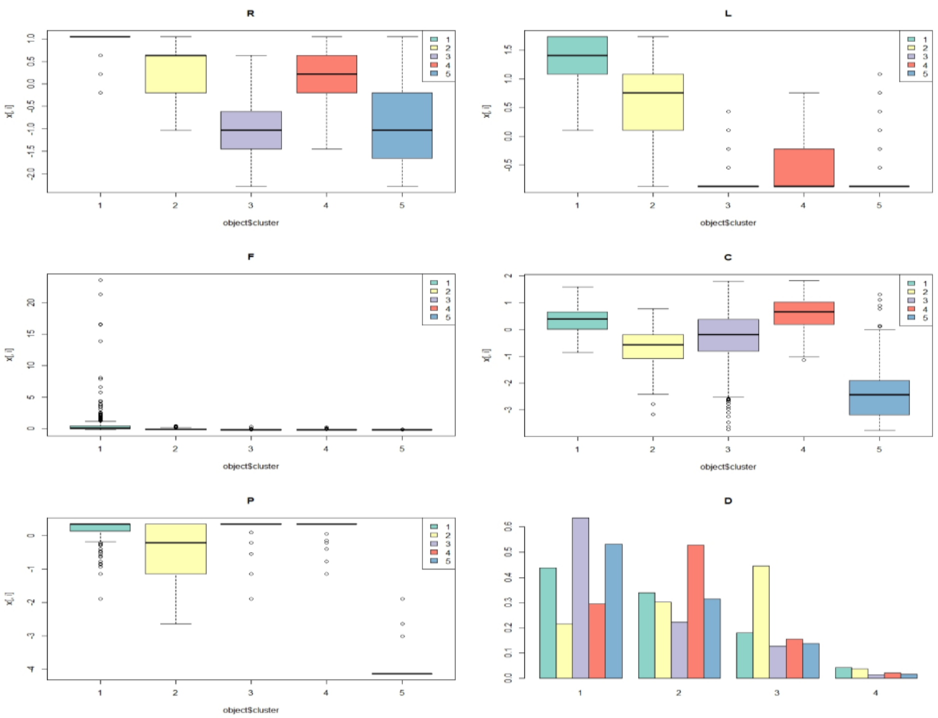



Fig. 11: The result of the K-prototype clustering

Table 4: The RLFCP-D charactersitics of five feeder schools clusters

| Cluster | N-R | N-L | N-F | N-C | N-P | Distribution of feeder schools in each D category | Number of schools | Number of students |
|---|---|---|---|---|---|---|---|---|
| C1 | 0.91 | 1.35 | 0.49 | 0.34 | 0.20 | 275;213;113;27 | 628 | 15025 |
| C2 | 0.37 | 0.53 | -0.10 | -0.66 | -0.42 | 53;74;109;9 | 245 | 1164 |
| C3 | -1.10 | -0.74 | -0.20 | -0.26 | 0.32 | 424; 149; 85;9 | 667 | 958 |
| C4 | 0.26 | -0.59 | -0.19 | 0.61 | 0.34 | 198;354;104;14 | 670 | 1240 |
| C5 | -0.86 | -0.76 | -0.21 | -2.31 | -3.80 | 66;39;17;2 | 124 | 150 |

N-* = average value of normalized * variable

As shown in Fig. 11, boxplots display the characteristics of the RLFCP variables. Cluster C1 has the highest median value for R, L, and F variables, while Cluster C4 has the highest median value for C and P variables. It means that based on the relationship values (R, L variables) and quantity value (F variable), C1 is the most valuable, while C4 has the most valuable quality value (C, P variables). For the categorical variable D, a bar chart depicts the distribution of the five clusters within each district's potential category. Most Cluster C1, C3, and C5 members are in D1 (the lowest potential). Most Cluster C2 members are at D3 (good potential), while most Cluster C4 members are at D2 (moderate potential).

According to Table 4, the feeder schools were classified into five clusters based on the RLFCP-D model. The cluster profiles are described as follows:

- C1: the most valuable feeder school cluster

C1 has the most valuable feeder schools, indicated by the positive value of all of the normalized RLFCP variables. The feeder schools have had long-standing relationships with the university and have sent many high-quality students. The cluster consists of 628 feeder schools, of which 275 feeder schools are in the lowest potential districts, 213 in moderate potential districts, 113 in good potential districts, and 27 in the best potential districts. This cluster's feeder schools sent the most students (15,025).

- C2: the low-performance feeder school cluster

C2 is made up of 245 feeder schools. Despite having a long-standing partnership until recently, they only sent a small number of students of low quality. However, most of the schools are in high-potential districts.

- C3: the low-value feeder school cluster

Although their alumni's persistence was high, most feeder schools in C3 are one-time feeder schools. The schools had not sent students for a long time, and their relationship was very short, so their contribution to the university was small. Most of the schools are in the lowest potential districts.

- C4: the highest-quality students' source cluster

C4 has feeder schools with the highest quality students but is still relatively new in establishing relationships with the university, expressed by the positive R-value but negative L, so not many students have been sent. Most of the feeder schools are in moderate potential districts.

- C5: the lowest-value feeder school cluster

C5 has the lowest-value feeder school because all of its RLFCP values are negative. Most of the feeder schools are in the lowest potential districts.

Since C1 is the most valuable cluster, it was chosen as the target segment. Although it only consists of 26.90% feeder schools, it provides 81.05% of students.

## 3.4. Temporal analysis for the highest value cluster

This research performed time-series clustering on the C1 cluster members, the most valuable feeder school cluster, to identify the temporal pattern of the number of enrolled students at the university. According to the Silhouette index, k=3 was chosen as the optimal number of clusters, with a value of 0.6621. Hence the time-series enrollment data were grouped into three patterns. The trend of the average number of enrolled students from the three time-series clusters is depicted in Fig. 12, and the characteristics of each cluster are presented in Table 5.

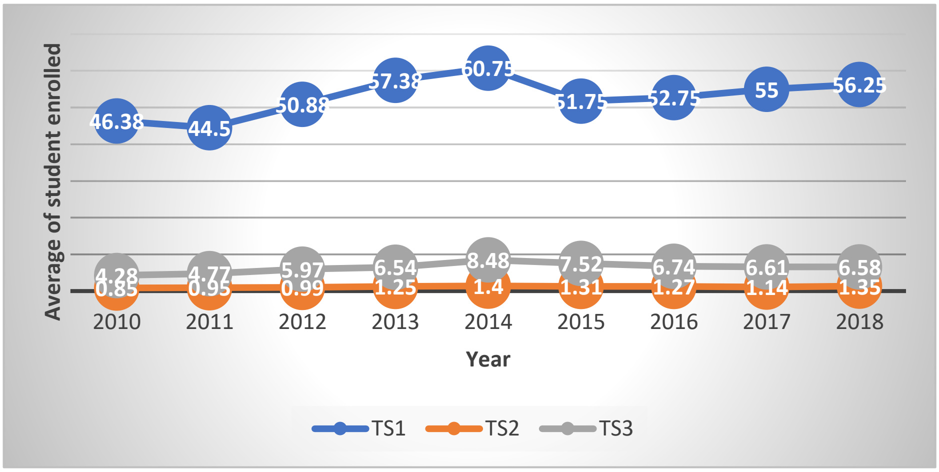



Fig. 12: The enrollment pattern of feeder schools in the highest value cluster

## 3.5. Feeder schools profiling

According to the time-series enrollment patterns, the target segment feeder schools have three patterns. Referring to Table 5, the following are their profiles:

**TS1** comprises eight feeder schools that have made long and very significant contributions to the university. The schools have the largest number of enrolled students every year, with a slightly increasing trend over the past few years after experiencing a decline, as shown in Fig. 12. The feeder schools produce high-quality students, most of them in the best potential districts.

**TS2** comprises 520 feeder schools that constantly send about one good-quality student per year. Most schools are located in the lowest potential districts.

**TS3** consists of 100 feeder schools that have had a long relationship with the university until recent times. The feeder schools send some good-quality students every year, although, with a slightly declining trend in the last few years, most of them are in moderate potential districts.

Table 5: The RLFCP-D statistics of time series clustering results

| Cluster | R | L | F | C | P | Distribution of feeder schools in each D category | Numbe r of schools | Num ber of stude nts |
|---|---|---|---|---|---|---|---|---|
| TS1 | 9 | 8 | 475.62 | 2.99 | 97.24 | 0;1;2;5 | 8 | 3805 |
| TS2 | 8.59 | 6.58 | 10.52 | 2.90 | 96.85 | 260;166;84;10 | 520 | 5471 |
| TS3 | 8.99 | 7.97 | 57.49 | 2.92 | 96.02 | 15;46;27;12 | 100 | 5749 |

This study recommends feeder schools in TS1 and TS3 as the priority target market for promotion, with TS1 being the most prioritized target feeder schools cluster. Fig. 13 depicts the distribution of the 108 prioritized target feeder schools.

The prioritized target schools are 4.63% of all feeder schools. However, they contributed to the university by sending 51.54% of enrolled students. The schools are spread out in 23 provinces, within 32 cities, and 23 regencies, consisting of 14 lowest potential districts, 28 moderate potential districts, 12 good potential districts, and one the best potential district. Based on the school's ownership, 71 feeder schools are private contributing 82% of students, and 37 public schools account for 18% of students. According to the inset figure in Fig. 13, most priority schools are located close to the university. Based on the findings, the university decision-makers can use them to develop a geo-marketing strategy.

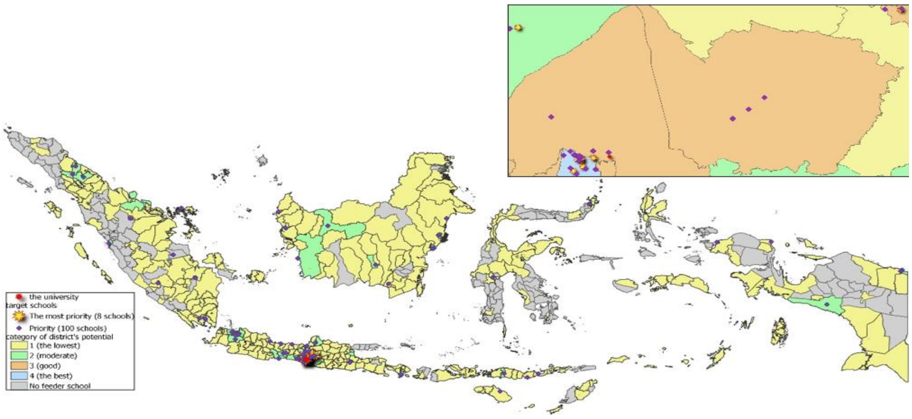
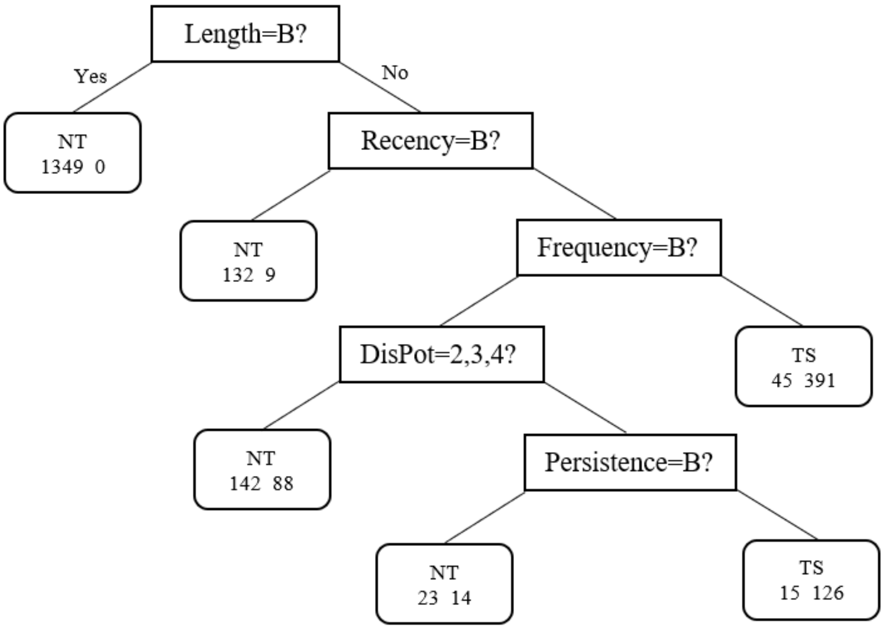



Fig. 13: The prioritized targeted feeder schools' distribution

## 3.6. Validation

The classification technique was used to construct a predictive decision tree model based on the clustering results. The decision tree generated based on the training data is shown in Fig.14, and it is then used for validation. Based on the 10-fold cross-validation applied ten times, the average accuracy, precision, recall, and F1-score values are 0.93, 0.91, 0.89, and 0.90. Compared to the RLFCP model (Ernawati et al., 2022a), precision, recall, and F1-score values of the RLFCP-D model were greater than the RLFCP model,  while for accuracy, both models performed similarly.

Fig. 14: The predictive decision tree model




## 3.7. Evaluation and recommendation

According to the decision tree, the target segment has characteristics: (Length above-average AND Recency above-average AND Frequency above-average) OR (Length above-average AND Recency above-average AND Frequency below or same to the average AND Persistence above-average). The decision tree rules can predict the next academic year's feeder schools' segment, as shown in Fig. 3. This study compared the predicted segment with the real class label obtained from the university's promotion office using 226 schools sample whose alumni were enrolled at the university in the following year. The accuracy, precision, recall, and F1-score values obtained were 0.85, 0.84, 0.86, and 0.85. Based on the testing conducted, the  RLFCP-D model is good enough for predicting the target segment. Besides being able to determine the target market based on the relationship, quantity, and quality values, the advantages of the RLFCP-D model are providing additional information about the feeder school district potential and considering the temporal pattern of the number of enrolled students, allowing this model more precise in choosing priority feeder schools from the target segment. Thus, the model can be used to assist in revealing the target segment and support decision-makers in developing geo-marketing strategies for promotion activities. However, this study has limitations. Since this study selected the priority target segment based on the cluster profiles, it cannot specify a predefined exact number of target schools for creating a list of feeder school rankings as a guide for making decisions regarding promotion implementation. Other MCDM methods, such as Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Simple Additive Weighting (SAW), or Analytic Hierarchy Process (AHP), can be added to rank the value of each priority feeder school and choose the n-best feeder schools as the target market.

Based on the findings of the case study, this study proposes the following recommendations to the university for their promotion activities: promotion should focus more on the feeder schools included in the target segment; perform intensive communication with the most priority target schools to increase engagement; prioritize promotion to private schools over public schools; and conduct school visits with on-site recruitment to priority schools in higher potential districts, where many feeder schools and students can be reached. Simultaneously, schools in low or moderate potential districts can be reached via email, by sending university students or alumni from the schools, or by participating in educational fairs.

# 4. Conclusion

This research combined spatial and temporal analysis with the modified RFM model to identify promotional target markets. GIS was used in the spatial analysis to determine the concentration of feeder schools and the potential of the feeder school's regencies/cities as a university target market. In the temporal analysis, time-series clustering revealed the temporal pattern of the number of students enrolled at the university from the feeder schools in the highest value segment. Based on a case study at an Indonesian university, this study proposes 108 priority feeder schools as the targeted schools using the RLFCP-D model and the K-prototype algorithm. Despite only 4.63% of all feeder schools, they significantly contributed to the university by sending 51.54% of enrolled students. Additionally, the proposed method can reveal the distribution of feeder schools in each district and the trend of enrolled students. Hence, it can assist university decision-makers in developing a geo-marketing strategy for promotional activities. The proposed model is a good predictor based on 10-fold cross-validation and testing conducted, with an accuracy of 0.93 and 0.85, respectively. This research can be developed to select n-best feeder schools from the target segment by incorporating an MCDM method, such as TOPSIS, SAW, or AHP.

# Acknowledgments

This paper is part of first author's Ph.D. research at the University Teknikal Malaysia Melaka, which is supported by Universitas Atma Jaya Yogyakarta.

# References

Abbasimehr, H. & Shabani, M. (2019). A new methodology for customer behavior analysis using time series clustering: a case study on a bank’s customers. *Kybernetes*

*50*(2), 221– 242.

Abbasimehr, H. & Shabani, M. (2021). A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques. *Journal of ambient intelligence and humanized computing*, *12*(1), 515-531.

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information Systems*, *53*, 16-38.

Aghdaie, M. H., & Alimardani, M. (2015). Target market selection based on market segment evaluation: a multiple attribute decision making approach. *International Journal of Operational Research*, *24*(3), 262-278.

Baviera-Puig, A., Buitrago-Vera, J., & Escriba-Perez, C. (2016). Geomarketing models in supermarket location strategies. *Journal of Business Economics and Management*, *17*(6), 1205-1221.

Beheshtian-Ardakani, A., Fathian, M., & Gholamian, M. (2018). A novel model for product bundling and direct marketing in e-commerce based on market segmentation. *Decision Science Letters*, *7*(1), 39-54.

BPS-Statistics Indonesia. (2019). *Statistical Yearbook of Indonesia 2019*. BPS-Statistics, Indonesia.

Bunnak, P., Thammaboosadee, S., & Kiattisin, S. (2015). Applying data mining techniques and extended RFM model in customer loyalty measurement. *Journal of Advances in Information Technology*, *6*(4), 238–242.

Chacón-García, J. (2017). Geomarketing techniques to locate retail companies in regulated markets. *Australasian Marketing Journal*, *25*(3), 185-193.

Chen, Y. S., Cheng, C. H., Lai, C. J., Hsu, C. Y., & Syu, H. J. (2012). Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment. *Computers in biology and medicine*, *42*(2), 213-221.

Cheng, Y., & Morimoto, Y. (2019). Analysis of supermarket customer behavior by using RFM and Huff 's gravity model. *Bulletin of Networking, Computing, Systems, and Software*, *8*(2), 81–86.

Chiang, W. Y. (2017). Discovering customer value for marketing systems: an empirical case study. *International Journal of Production Research*, *55*(17), 5157-5167.

Cliquet, G., & Baray, J. (2020). *Location-based Marketing: Geomarketing and Geolocation*. ISTE Ltd, London & John Wiley & Sons, Inc, Hoboken, NJ.

Constantinidis, B. R. (2019). Geomarketing as a location public - private decision support tool or schools in the City of Buenos Aires, Argentina. *Territorio Italia, 1,2,* 31–54. https://doi.org/10.14609/Ti

Dahiya, A., Gautam, N., & Gautam, P. K. (2021). Data mining methods and techniques for online customer review analysis: a literature review. Journal of System and Management Sciences, 11(3), 1–26.

Ernawati, Baharin, S. S. K., & Kasmin, F. (2021). Spatial-temporal analysis using two-stage clustering and GIS-based MCDM to identify potential market regions. *Journal of System and Management Sciences*, *11*(4), 87–112.

Ernawati, Baharin, S. S. K., & Kasmin, F. (2022a). Modified RFM model and the FUCOM method for identifying higher education institution target market. *ICIC Express Letters, in press*.

Ernawati, Baharin, S. S. K., & Kasmin, F. (2022b). Target Market Determination for Information Distribution and Student Recruitment Using an Extended RFM Model with Spatial Analysis. *The Journal of Distribution Science, in press*.

Ghodousi, M., Sadeghi-Niaraki, A., Rabiee, F., & Choi, S. M. (2020). Spatial-temporal analysis of point distribution pattern of schools using spatial autocorrelation indices in Bojnourd city. *Sustainability*, *12*(18), 7755.

Guarda, T., & Augusto, M. F. (2019). Geographic marketing intelligence: GMI model. *Second International Conference on Applied Informatics (ICAI Workshop)*, 205-214.

Hidayat, S., Rismayati, R., Tajuddin, M., & Merawati, N. L. P. (2020). Segmentation of university customers loyalty based on RFM analysis using fuzzy c-means clustering. *Jurnal Teknologi dan Sistem Komputer*, *8*(2), 133-139.

Hosseini, M., & Shabani, M. (2015). New approach to customer segmentation based on changes in customer value. *Journal of Marketing Analytics*, *3*(3), 110-121.

Hosseini, Z. Z., & Mohammadzadeh, M. (2016). Knowledge discovery from patients' behavior via clustering-classification algorithms based on weighted eRFM and CLV model: an empirical study in public health care services. *Iranian journal of pharmaceutical research*, *15*(1), 355-367.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, *2*(3), 283-304.

Hwang, S., & Lee, Y. (2021). Identifying customer priority for new products in target marketing: Using RFM model and TextRank. *Innovative Marketing*, *17*(2), 125-136.

Javed, A., Lee, B. S., & Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications*, *1*, 100001.

Kelly, M. G. (2019). A map is more than just a graph: Geospatial educational research and the importance of historical context. *AERA Open*, *5*(1), 2332858419833346.

Kit, T. C., Firdaus, N., & Azmi, M. (2021). Customer profiling for Malaysia online retail industry using K-Means clustering and RM model. *International Journal of Advanced Computer Science and Applications*, *12*(1), 106–113.

Lamb, C. W., Hair, J. F. J., & McDaniel, C. (2011). *Marketing*. South-Western Cengage Learning, Mason, OH.

Lawson, R. G., & Jurs, P. C. (1990). New index for clustering tendency and its application to chemical problems. *Journal of chemical information and computer sciences*, *30*(1), 36-41.

Levin, N., & Zahavi, J. (2005). Data mining for target marketing. In *Data mining and knowledge discovery handbook* (pp. 1261-1301). Springer, Boston, MA

Li, T., Wu, X., & Zhang, J. (2020). Time series clustering model based on DTW for classifying car parks. *Algorithms*, *13*(3), 57.

Libório, M. P., Bernardes, P., Ekel, P. I., Ramalho, F. D., & dos Santos, A. C. G. (2020). Geomarketing and the locational problem question in the marketing studies. *REMark*, *19*(2), 448-469.

Mekvabidze, R. (2020). From business modeling to business management: an exploratory study of the optimal decision making on the modern university level. Journal of Logistics, Informatics and Service Science, 7(1), 67–86.

Melnyk, L., & Nyzhnyk, L. (2018). Geomarketing is an innovative technology business. *Industry 4.0*, *3*(3), 141-143.

Morris, P., & Thrall, G. (2010). Using geospatal technique to address institutional objective: St. Petersburg College geo-demographic analysis. IR Applications, Volume 27. *Association for Institutional Research (NJ1)*.

Nunes, A., Santana, C., Bezerra, F., & Sobral, N. (2014). Knowledge acquisition based on geomarketing information for decision making: a case study on a food company. *International Journal of Innovation, Management and Technology*, *5*(6), 422-427.

Pardiyono, R., J. Suteja, H. Puspita, and U. Juju. (2022). Dominant factors for the marketing of private higher education. *Decision Science Letters, 11*(2), 137-146.

Suhaibah, A., Uznir, U., Rahman, A. A., Anton, F., & Mioc, D. (2016). 3D geomarketing segmentation: a higher spatial dimension planning perspective. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*.

Sukoroto, Haryono, S., & Kharisma, B. (2020). Target market selection using MCDM approach: a study of rolling stock manufacturer. *The Journal of Distribution Science, 18*(7), 63-72.

Sulastri, S., Usman, L., & Syafitri, U. D. (2021). K-prototypes algorithm for clustering schools based on the student admission data in IPB university. *Indonesian Journal of Statistics and Its Applications*, *5*(2), 228-242.

Surjandari, I., and Rosyidah, A. (2017). Fixed broadband customer area mapping using spatial analysis. 2017 IEEE 8th International Conference on Awareness Science and Technology, 109–114.

Verschuren, M. M. J. (2006). *Geomarketing. GIS & marketing, the new combination of knowledge* Master's thesis, Wageningen University.

Vitor, J. A., Guarda, T., Augusto, M. F., Leon, M., Villao, D., Mazon, L., & Estrada, Y. S. (2018). Inexpensive marketing tools for SMEs. *International Conference on Information Technology & Systems*, 129-137.