ICIC Express Letters
Volume **16**, Number **9**, September **2022**

ICIC International ©2022 ISSN 1881-803X
pp. 993–1000

MODIFIED RFM MODEL AND THE FUCOM METHOD FOR IDENTIFYING HIGHER EDUCATION INSTITUTION TARGET MARKET

Ernawati[1,*], Safiza Suhana Kamal Baharin[2] and Fauziah Kasmin[2]

[1]Department of Informatics
Universitas Atma Jaya Yogyakarta
Jalan Babarsari 43, Yogyakarta 55281, Indonesia
*Corresponding author: ernawati@uajy.ac.id

[2]Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya, Durian Tunggal 76100, Melaka, Malaysia
{ safiza; fauziah }@utem.edu.my

Received August 2021; accepted November 2021

Abstract. *Promotion to the right target market helps higher education institutions (HEIs) in increasing student recruitment success. The most common method to identify target customers is the recency, frequency, and monetary (RFM) model combined with data mining techniques. The RFM model measures customer value by business profitability. However, the educational institution has different profit values. Therefore, to determine the feeder school's value for the target market, this study modifies the RFM model by considering HEI characteristics. The proposed model assesses feeder schools from three perspectives: relationship, quantity, and quality. K-means algorithm and full consistency method (FUCOM) are used to reveal feeder schools characteristics for target segment determination, and the classification and regression tree (CART) is used to evaluate the predictive model's performance. By examining 2,334 schools with 20,469 graduates enrolled at one Indonesian private university, this study identified ten feeder school groups and set four clusters with the highest value as the target segment. The target segment includes 15.12% of high-quality feeder schools that consistently supply students. The proposed model can predict the target market with 93% accuracy. This finding shows that the proposed model successfully identifies HEI's target market to help develop an effective promotional strategy.*
Keywords: Data mining, FUCOM, Higher education, RFM model, Target market

1. **Introduction.** The private education market is becoming more competitive. Enrollment shifts from private to public schools may occur due to the COVID-19 pandemic [1]. In this competitive market, developing a marketing strategy that includes promotions is critical. A targeted marketing campaign can increase recruitment success and save resources. Most universities recruit students from senior high school graduates, implying high schools serve as feeder schools. Identifying appropriate feeder schools aids higher education institutions (HEIs) in recruiting prospective students.

With 3,044 private HEIs in Indonesia, competition is fierce. Fortunately, 27,930 senior high schools and vocational high schools [2] could be a potential market. However, promotion to many schools requires many resources. So, reaching all high schools is impossible. Moreover, not all schools are equally valuable for the university. Therefore, determining which high schools should be targeted for promotion becomes critical.

Customers' value is the basic information for determining the target market. The recency, frequency, monetary (RFM model) is commonly used to estimate customer value [3]

DOI: 10.24507/icicel.16.09.993

and target market [4]. The RFM model specifies the business profit value using recency, frequency, and monetary. Recency refers to the novelty of the latest purchase. Frequency measures the number of purchases, and monetary refers to the total money paid [5]. More valuable customers have a higher frequency and monetary values and more recent transactions. However, HEI has different profit values than businesses organizations.

The previous study used data mining to identify the target market. C4.5 algorithm was used in [6] to categorize schools for university marketing based on demographic attributes. RFM-based model is another popular way to target a market. [4] proposed the RF and RwF (recency, frequency, and utility-weighted frequency) model to find prospective customers for new products. Most previous research combined the RFM model and data mining techniques to identify customers. In [7], the RFM model and K-means algorithm classified customers based on purchase behaviors, while in [8], the RFM+B model, which added customer balances to the RFM model and the K-means algorithm, was employed to segment banking customers. The time, frequency, monetary (TFM) score was proposed in [9] to segment and target telecom customers. They used classification algorithms to discover the causes and influential attributes to loyalty and categorize new users. [10] used Fuzzy C-means and redefined the RFM model to segment high school loyalty to the university. They redefined recency as the number of times a specific school's alumni registered in the university and frequency as average registrants from the school every year. They included the monetary variable, but it refers to the total registrant from a school. Although the RFM model has succeeded in identifying the high schools' loyalty, the critical value for HEI has not been captured. Therefore, this study aims to modify the RFM model by considering HEI's values as a non-profit organization and service provider [11].

Relationship marketing is appropriate for HEI as a service provider [12]. Maintaining a good relationship with feeder schools will increase their trust and loyalty to the university. While as a non-profit organization, HEI marketing goals are not solely focused on financial gains [12]. So, this study modifies the RFM model by replacing the monetary variable with the relationship length. The length of the relationship variable is widely used [13-15] because it is closely related to customer loyalty [14]. The longer the relationship is maintained, the more committed the feeder school is. The academic success of a feeder school's alumni is also important. The most widely used academic achievement metric is the grade point average (GPA). Most Indonesian universities evaluate students at the end of the fourth semester based on their cumulative GPA (CGPA). Another study used graduation persistence as the primary indicator of student success [16]. Non-persistent students drop out or withdraw from their studies, often in their first or second year [17]. Dropouts have consequences for both the student and the institution [18]. It reduces revenue, graduation rates, and university reputation [19]. Therefore, the proposed model includes recency, length, frequency, CGPA, persistence (RLFCP) variables for identifying feeder school's value for the HEI target market.

This research examines real data from one Indonesian private university to validate the proposed approach. To determine the target segment, the RLFCP model is combined with the K-means algorithm and full consistency method (FUCOM), while the classification and regression tree (CART) algorithm is used to evaluate the predictive model's performance. The 10-fold cross-validation is performed using accuracy, precision, recall, and F1-score metrics. The contributions of this study are summarized as follows: 1) Provide a novel model for quantifying feeder school value by adopting the characteristics of an educational institution as a non-profit service provider organization, so the proposed RLFCP model can assess feeder schools from three perspectives: relationship, quantity, and quality; 2) Propose a new framework using the RLFCP model for determining the promotional target market; 3) Provide an empirical case that integrates the RFM based model and FUCOM method to determine feeder school value and the targeted promotional segment. The rest of this paper is structured as follows. Section 2 introduces the proposed RLFCP model, the basic theory of the FUCOM method, and the proposed framework, while the empirical finding and analysis are presented in Section 3. The last section summarizes a brief conclusion.

2. Research Method.

2.1. **RLFCP model.** The proposed RLFCP model for determining the feeder school's value includes five variables that reflect relationship, quantity, and quality, as shown in Table 1. The recency and length indicate the strength of the feeder school's relationship with the university. The frequency quantifies the feeder school's supply to the university. The CGPA and persistence are used to assess the quality of feeder school graduates.

Table 1. The RLFCP model variables and definition

Value perspective	Variable	Definition
Relationship	Recency (R)	The novelty of a feeder school's last alumnus enrolled at the university in the analysis period, on a 1-9 scale.
	Length (L)	The length of time since the first and last alumni of a feeder school enrolled at the university, on a 0-8 scale.
Quantity	Frequency (F)	The total number of alumni from a feeder school enrolled at the university during the analysis period.
Quality	CGPA (C)	The average CGPA of a feeder school's alumni at the university's end of the fourth semester, on a 0-4 scale.
	Persistence (P)	The percentage of a feeder school's alumni who do not drop out or withdraw until the end of the fourth semester.

2.2. **FUCOM method.** This method is based on pairwise comparisons, while results are validated by deviation from full consistency (DFC). FUCOM requires fewer pairwise comparisons than the analytical hierarchy process (AHP) [20,21], which is one of the most commonly used MCDM methods for ranking customer value but suffers from an exponential increase in comparisons [21]. The FUCOM algorithm has three steps [20].
1) Ranking the variable based on decision-maker judgment, starting with the variable expected to have the highest importance:

$$V_1 > V_2 > \cdots > V_k > \cdots > V_m \tag{1}$$

2) Determining the comparative priority $\left(\varphi_{k/(k+1)}\right)$ by comparing the V_k rank scale to the V_{k+1} in Equation (1). One way to find the comparative priority vector $\Phi = \left(\varphi_{1/2}, \varphi_{2/3}, \ldots, \varphi_{k/(k+1)}, \ldots, \varphi_{(m-1)/m}\right)$ is by determining the priority of each variable (ω_{V_k}) concerning the most significant variable and make an $n-1$ comparison to them.
3) Calculate the weight coefficients $(w_1, w_2, \ldots, w_m)$ and DFC (χ). The weight coefficients should meet the two requirements:

(a) the weight coefficients' ratio is equal to the comparative priority among the observed variables $\left(\varphi_{k/(k+1)}\right)$:

$$\frac{w_k}{w_{k+1}} = \varphi_{k/(k+1)} \tag{2}$$

(b) satisfy the mathematical transitivity restrictions:

$$\frac{w_k}{w_{k+2}} = \varphi_{k/(k+1)} \otimes \varphi_{(k+1)/(k+2)} \tag{3}$$

To get the weight coefficients, solve the model in Equation (4). It seeks the minimum χ that meets the requirements in step 3(a) (Equation (4) line 1), step 3(b) (line 2), and each weight greater than zero, with the sum of all variable weights equal to one (line 3).

$$\text{Min}\chi \quad \text{s.t.} \begin{cases} \left| \dfrac{w_k}{w_{k+1}} - \varphi_{k/(k+1)} \right| \leq \chi, \quad \forall k \\ \left| \dfrac{w_k}{w_{k+2}} - \varphi_{k/(k+1)} \otimes \varphi_{(k+1)/(k+2)} \right| \leq \chi, \quad \forall k \\ \sum w_k = 1, \quad w_k \geq 0, \quad \forall k \end{cases} \tag{4}$$

2.3. **The proposed framework.** Figure 1 shows the proposed approach in three phases: 1) data preprocessing, 2) segmentation and profiling, and 3) target segment determination and evaluation. Data preprocessing involves integrating, selecting, and cleaning data.

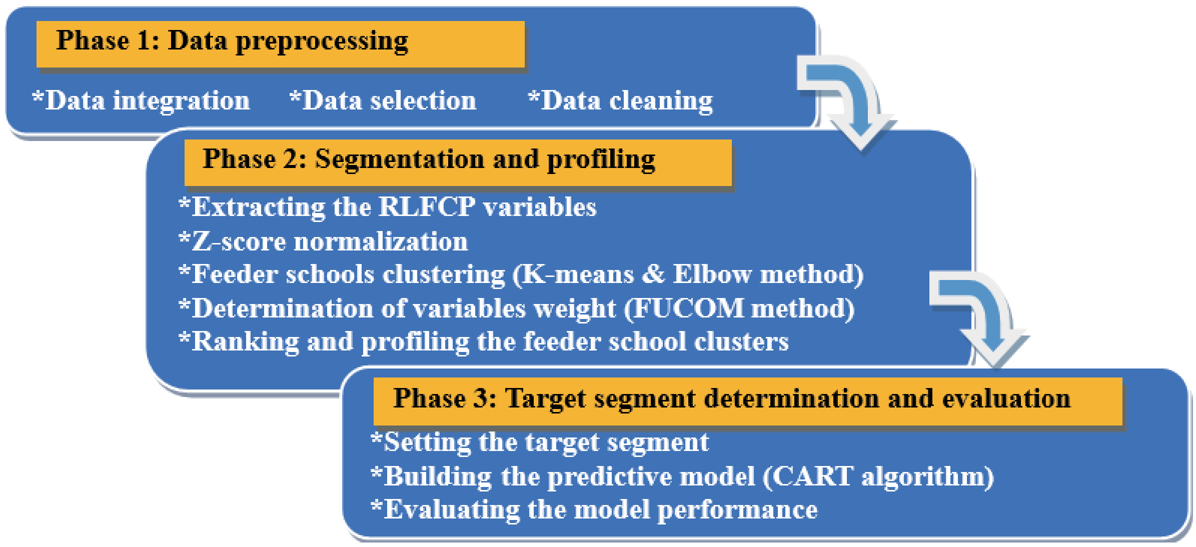

FIGURE 1. The proposed framework

Phase 2 starts with extracting RLFCP data for each feeder school. Before clustering, the data is normalized. We chose Z-score normalization to make it easier to interpret feeder school characteristics. A positive Z value indicates the real value is greater than the sample's mean; otherwise, it is lower. Clustering is done with the K-means algorithm and the Elbow method for determining the optimum k value [7]. The cluster's score is calculated using weighted-centroid in Equation (5), then ranked and profiled.

$$FV_{C_j} = W_R \cdot NR_{C_j} + W_L \cdot NL_{C_j} + W_F \cdot NF_{C_j} + W_C \cdot NC_{C_j} + W_P \cdot NP_{C_j} \tag{5}$$

NR_{C_j}, NL_{C_j}, NF_{C_j}, NC_{C_j}, NP_{C_j} in Equation (5) are the normalized centroid of the RLFCP variables for cluster C_j. Whereas W_R, W_L, W_F, W_C, W_P are the relative weights of the RLFCP variables obtained by the FUCOM method.

In the last phase, the feeder school target segment is determined. After discretizing the target segment features below and above average, a predictive model is built [13] using a classification algorithm to evaluate the proposed model [22]. The performance evaluation is carried out by 10-fold cross-validation using the accuracy, precision, recall, and F1-score metrics. For analysis, the RLFCP model results are compared to recency-frequency (RF), recency-frequency-CGPA (RFC), and recency-length-frequency-CGPA (RLFC) models.

3. Empirical Analysis.

3.1. **Data preprocessing.** This study uses registration and academic data from one Indonesian private university. After integrating and cleaning data, the final dataset contains 18,537 student records from 20,469 students enrolled in 2010 to 2018. Aggregating students to their feeder school, this study includes 2,334 feeder schools for analysis.

3.2. **Feeder schools segmentation and profiling.** Table 2 summarizes RLFCP statistics. The large standard deviation values show a large variability between schools' features.

Clustering using K-means and Elbow method produced five clusters. All RLFCP variables in C2 and C4 cluster centroids are positive, indicating that the feeder schools in both clusters are valuable for the target market. However, due to the frequency, CGPA, persistence values approach 0 for C4, and it has many schools (783), it requires more thorough consideration using two-stage clustering [15]. So, C4 is regrouped into six new clusters, giving a total of ten clusters. Table 3 shows the normalized clusters' centroid (NR, NL, NF, NC, NP) and its real value (R, L, F, C, P), the number of schools and students involved in each cluster, while Figure 2 visualized the clusters' characteristics by their normalized centroid. Positive values for all attributes indicate that schools in clusters C2, C4.1, C4.5, and C4.6 are worth being the target market for promotion.

Table 2. RLFCP summary statistics

Variable	Minimum	Maximum	Mean	Standard deviation
Recency	1	9	6.47	2.40
Length	0	8	2.68	3.07
Frequency	1	780	7.94	32.70
CGPA	0.03	3.95	2.67	0.70
Persistence	0	100	92.17	22.26

Table 3. The clustering results

Cluster	Cluster centroid										School		Student	
	NR	R	NL	L	NF	F	NC	C	NP	P	N	%	N	%
C1	0.30	7.20	−0.59	0.88	−0.19	1.83	0.57	3.07	0.33	99.42	729	31.23	1335	7.20
C2	1.05	9.00	1.74	8.00	18.37	608.80	0.54	3.05	0.23	97.34	5	0.21	3044	16.42
C3	−0.79	4.57	−0.69	0.55	−0.20	1.31	−2.21	1.12	−3.64	11.09	136	5.83	178	0.96
C4.1	1.05	8.98	1.72	7.96	1.38	52.98	0.37	2.93	0.18	96.11	89	3.81	4715	25.44
C4.2	0.70	8.15	1.18	6.30	0.05	9.60	−0.44	2.36	0.12	94.76	178	7.63	1709	9.22
C4.3	0.76	8.28	0.81	5.16	−0.06	5.83	0.50	3.02	0.27	98.16	190	8.14	1108	5.98
C4.4	0.45	7.55	0.87	5.33	−0.10	4.76	−0.66	2.20	−1.17	66.22	67	2.87	319	1.72
C4.5	1.05	9.00	1.74	8.00	5.14	176.20	0.37	2.93	0.22	97.05	10	0.43	1762	9.50
C4.6	1.00	8.87	1.60	7.59	0.17	13.36	0.47	3.00	0.23	97.25	249	10.67	3326	17.94
C5	−1.13	3.76	−0.70	0.52	−0.20	1.53	−0.36	2.41	0.28	98.31	681	29.18	1041	5.62

The FUCOM method [20] weighs each RLFCP variable according to the university's Partnership and Promotion officer preference:

1) The RLFCP variables were ranked as follows: F>L>R>C>P. It means the frequency is the most important, while the persistence is the least important.

2) Based on the variables' priorities (ω_{V_k}) in Table 4 from the decision-maker preference, we computed the comparative priorities: $\varphi_{F/L} = 2.5/1 = 2.5$; $\varphi_{L/R} = 4/2.5 = 1.6$; $\varphi_{R/C} = 5/4 = 1.25$; $\varphi_{C/P} = 6/5 = 1.2$, yielded the comparative priority vector $\Phi = (2.5, 1.6, 1.25, 1.2)$.

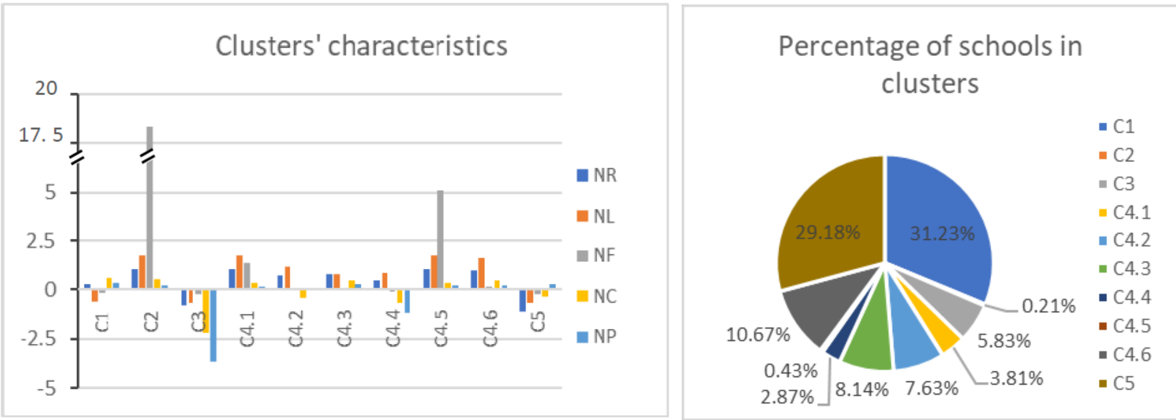

FIGURE 2. The clusters' characteristics

TABLE 4. The priorities of the RLFCP variables

Variable	F	L	R	C	P
ω_{V_k}	1	2.5	4	5	6

3) By considering Equations (2) and (3) conditions in Equation (4), we obtained Equation (6). The weight coefficients $(W_F, W_L, W_R, W_C, W_P) = (0.496, 0.198, 0.124, 0.099, 0.083)$ with $\text{DFC}(\chi) = 0$ were obtained after solving Equation (6).

$$\text{Min}\chi \text{ s.t.} \begin{cases} \left|\frac{W_F}{W_L} - 2.5\right| \leq \chi, \quad \left|\frac{W_L}{W_R} - 1.6\right| \leq \chi, \quad \left|\frac{W_R}{W_C} - 1.25\right| \leq \chi \\ \left|\frac{W_C}{W_P} - 1.2\right| \leq \chi, \quad \left|\frac{W_F}{W_R} - 4\right| \leq \chi, \quad \left|\frac{W_L}{W_C} - 2\right| \leq \chi, \quad \left|\frac{W_R}{W_P} - 1.5\right| \leq \chi \\ \sum W_k = 1, \quad W_k \geq 0, \quad \forall k \end{cases} \tag{6}$$

Using the FUCOM result, the cluster score and rank were calculated. Cluster profiles are obtained by analyzing the clusters' characteristics and are useful to identify the target segment. Figure 3 depicts the cluster score, rank, and profile summary.

3.3. **Target segment determination and evaluation.** Based on the cluster profiles, the decision-maker chose the first to fourth rank clusters as the target segment for their positive relationship, quantity, and quality values. The target segment comprises feeder schools with long-standing relationships that consistently supply high-quality graduates. This segment encompasses 15.12% of schools supplied 69.30% of students. The university can reach this segment by hosting an open house or school visit promotion by conducting on-site tests. It may use virtual meetings followed by an online test throughout this pandemic period. Table 5 compares the RLFCP model's performance and target segment features to the RF, RFC, and RLFC models. The RLFCP is a good model for predicting the target market since its accuracy, precision, and recall are greater than 0.75 [23], so do the F1-score, the geometric mean of recall and precision. RLFCP has slightly better performance than RFC and RLFC, but still below the RF model. However, the RLFCP model's target segment characteristics are comprehensively better than the RF model, except for frequency, because the RF model's clustering focuses more on recency and frequency attributes. As a result, using the RLFCP model enables finding a target segment with a better relationship, quantity, and quality values simultaneously, so the university can build relationships with loyal feeder schools and get sufficient good-quality students.

Figure 3. The clusters' rank and profile

Table 5. Comparison of model performance and target segment features

Model	Performance				Target segment features					School	Student
	Accuracy	Precision	Recall	F1-score	R	L	F	C	P	(%)	(%)
RLFCP	0.93	0.88	0.85	0.86	8.90	7.70	36.39	2.98	96.96	15.12	69.30
RLFC	0.93	0.87	0.84	0.86	8.91	7.72	35.95	2.96	96.71	15.51	70.21
RFC	0.88	0.84	0.80	0.82	8.95	6.12	26.47	2.90	96.43	22.75	75.82
RF	0.96	0.88	0.97	0.93	8.85	7.45	40.45	2.85	95.40	14.39	73.31

4. **Conclusion and Future Work.** For identifying high schools as HEI's target market, this study proposes the RLFCP model. Considering HEI as a non-profit service provider, the RLFCP model assesses feeder schools from three perspectives: relationship, quantity, and quality. Using real data from Indonesian private university, the RLFCP model combined with the FUCOM method and data mining techniques successfully identified four clusters with the highest value feeder schools as the target market. The RLFCP model can comprehensively identify the target market based on recency, length, frequency, CGPA, and persistence attributes. The proposed model's analytical findings can help HEI

management develop an effective promotional strategy and allocate resources for promotional activities. This study did not consider the school's location. Thus, we will expand it in the future with spatial analysis and spatial RFM. To improve the predictive model performance, we can look at other classification algorithms.

REFERENCES

[1] S. Al-Samarrai, M. Gangwar and P. Gala, *The Impact of the COVID-19 Pandemic on Education Financing*, World Bank Group, Washington, D.C., 2020.

[2] BPS-Statistics Indonesia, *Statistical Yearbook of Indonesia 2021*, BPS-Statistics Indonesia, Jakarta, 2020.

[3] W. Y. Chiang, Discovering customer value for marketing systems: An empirical case study, *Int. J. Prod. Res.*, vol.55, no.17, pp.5157-5167, 2017.

[4] S. Hwang and Y. Lee, Identifying customer priority for new products in target marketing: Using RFM model and TextRank, *Innov. Mark.*, vol.17, no.2, pp.125-136, 2021.

[5] E. Ernawati, S. S. K. Baharin and F. Kasmin, A review of data mining methods in RFM-based customer segmentation, *J. Phys. Conf. Ser.*, vol.1869, 012085, 2021.

[6] S. A. Abaya and B. D. Gerardo, An education data mining tool for marketing based on C4.5 classification technique, *The 2nd Int. Conf. on E-Learn. and E-Techn. in Educ.*, Lodz, Poland, pp.289-293, 2013.

[7] J. Wu et al., An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm, *Math. Probl. Eng.*, vol.2020, 2020.

[8] U. Firdaus and D. N. Utama, Development of bank's customer segmentation model based on RFM+B approach, *ICIC Express Letters, Part B: Applications*, vol.12, no.1, pp.17-26, 2021.

[9] W. N. Wassouf et al., Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study, *Journal of Big Data*, 2020.

[10] S. Hidayat et al., Segmentation of university customers loyalty based on RFM analysis using fuzzy C-means clustering, *J. Teknol. and Sist. Komput.*, vol.8, no.2, pp.133-139, 2020.

[11] L. Nicolescu, Applying marketing to higher education: Scope and limits, *Manag. Mark.*, vol.4, no.2, pp.35-44, 2009.

[12] J. Hemsley-Brown and I. Oplatka, Universities in a competitive global marketplace: A systematic review of the literature on higher education marketing, *Int. J. Publ. Sect. Man.*, vol.19, no.4, pp.316-338, 2006.

[13] Z. Z. Hosseini and M. Mohammadzadeh, Knowledge discovery from patients' behavior via clustering-classification algorithms based on weighted eRFM and CLV model: An empirical study in public health care services, *Iran. J. Pharm. Res.*, vol.15, no.1, pp.355-367, 2016.

[14] S. Peker, A. Kocyigit and P. E. Eren, LRFMP model for customer segmentation in the grocery retail industry: A case study, *Mark. Intell. Plan.*, vol.35, no.4, 2017.

[15] A. Sheikh, T. Ghanbarpour and D. Gholamiangonabadi, A case study of fintech industry: A two-stage clustering analysis for customer segmentation in the B2B setting, *Journal of Business-to-Business Marketing*, 2019.

[16] R. L. Altman, *Persistence and Achievement in Academics*, Master Thesis, Georgia Southern University, 2017.

[17] C. Ousley, *A Geographic-Information-Systems-Based Approach to Analysis of Characteristics Predicting Student Persistence and Graduation*, Ph.D. Thesis, The University of Arizona, 2010.

[18] L. Kemper, G. Vorhoff and B. U. Wigger, Predicting student dropout: A machine learning approach, *Eur. J. High. Educ.*, vol.10, no.1, pp.28-47, 2020.

[19] B. Perez, C. Castellanos and D. Correal, Applying data mining techniques to predict student dropout: A case study, *2018 IEEE 1st Colomb. Conf. on Appl. Comput. Intell.*, Medellin, Colombia, 2018.

[20] D. Pamučar, Ž. Stević and S. Sremac, A new model for determining weight coefficients of criteria in MCDM models: Full Consistency Method (FUCOM), *Symmetry*, vol.10, no.9, pp.1-22, 2018.

[21] M. O. Esangbedo and S. Bai, Grey regulatory focus theory weighting method for the multi-criteria decision-making problem in evaluating university reputation, *Symmetry*, vol.11, no.2, 2019.

[22] J. B. Pernabas and S. F. Fidele, Enhancements to greedy web proxy caching algorithms using data mining method and weight assignment policy, *International Journal of Innovative Computing, Information and Control*, vol.14, no.4, pp.1311-1326, 2018.

[23] T. Zimmermann et al., Cross-project defect prediction: A large scale experiment on data vs. domain vs. process, *Proc. of the 7th Joint Meeting of the European Soft. Engineering Conf. and the ACM SIGSOFT Int. Symposium on the Foundations of Soft Engineering*, Amsterdam, The Netherlands, pp.91-100, 2009.