



**MODIFIED RECENCY-FREQUENCY-MONETARY MODEL
FOR DETERMINING FEEDER SCHOOLS AS GEOMARKETING
PROMOTIONAL TARGET IN INDONESIA**

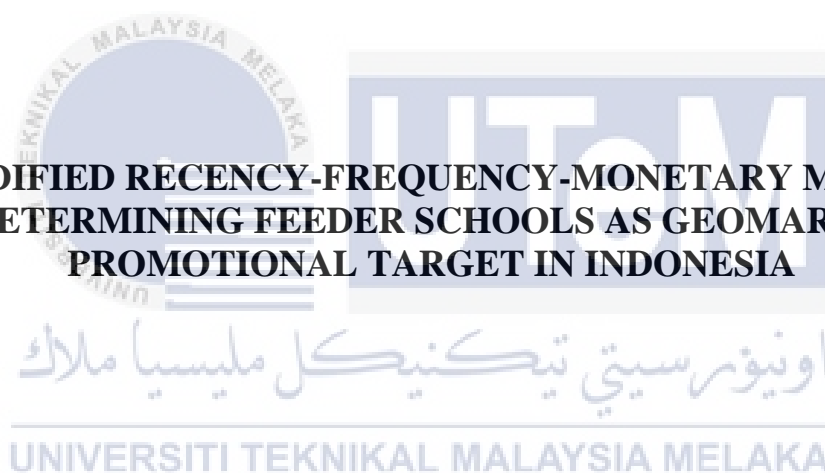


DOCTOR OF PHILOSOPHY

2023



Faculty of Information and Communication Technology



**MODIFIED RECENCY-FREQUENCY-MONETARY MODEL
FOR DETERMINING FEEDER SCHOOLS AS GEOMARKETING
PROMOTIONAL TARGET IN INDONESIA**

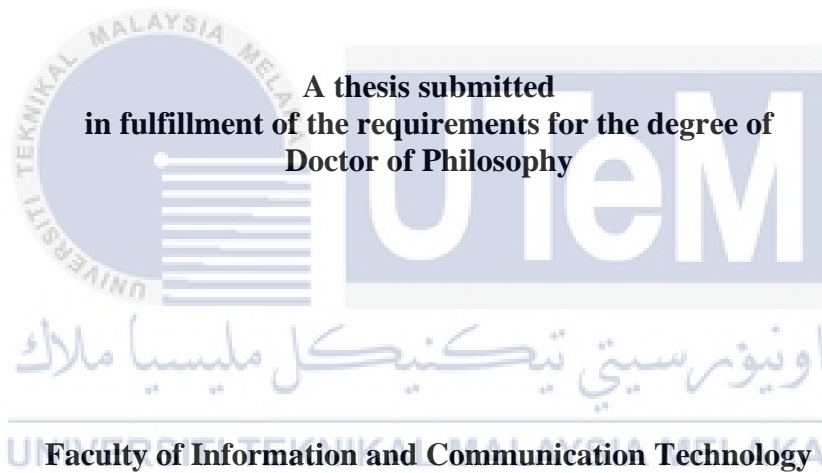
Ernawati

Doctor of Philosophy

2023

**MODIFIED RECENCY-FREQUENCY-MONETARY MODEL
FOR DETERMINING FEEDER SCHOOLS AS GEOMARKETING
PROMOTIONAL TARGET IN INDONESIA**

ERNAWATI



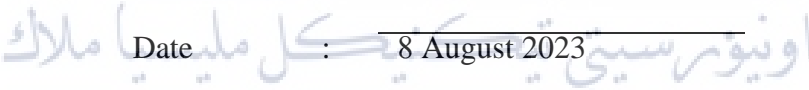


UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I declare that this thesis entitled “Modified Recency-Frequency-Monetary Model for Determining Feeder Schools as Geomarketing Promotional Target in Indonesia” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

 Signature : 
Name : Ernawati
Date : 8 August 2023 

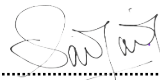
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of the Doctor of Philosophy.

Signature

:



Supervisor Name

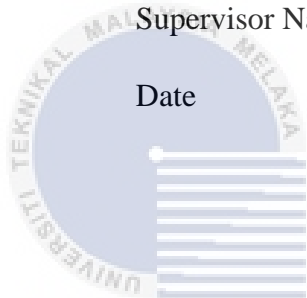
:

Gs. Dr. Safiza Suhana Kamal Baharin

Date

:

8 August 2023

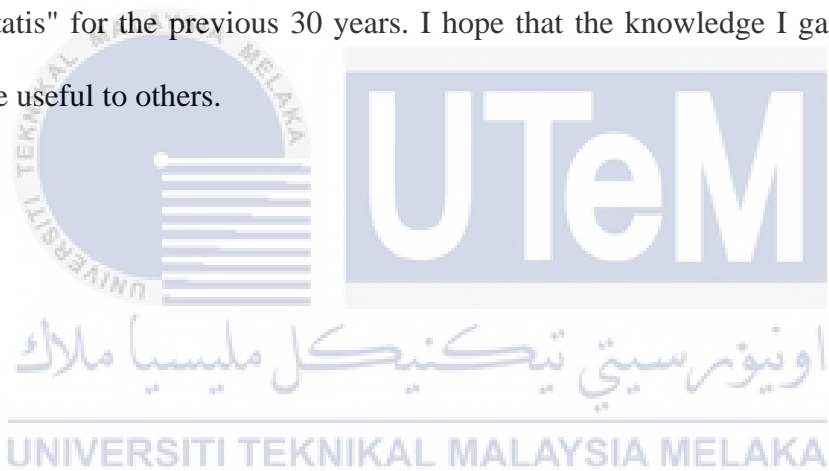


اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEDICATION

I dedicate this thesis to my Almighty God, for His grace enabled me to finish it. I also dedicate it to my loved ones, especially my husband and daughter, who are always there for me with prayers, support, and understanding. In addition, I also dedicate this thesis to Universitas Atma Jaya Yogyakarta, which has trusted me to serve through "serviens in lumine veritatis" for the previous 30 years. I hope that the knowledge I gained from this study will be useful to others.



ABSTRACT

Promotion to the right target market helps higher education institutions (HEIs) increase student recruitment success and save costs. High schools are HEI student feeders. Thus, promoting to the schools is critical. However, due to many feeder schools and their vast distribution, particularly in Indonesia, HEI management should choose high-value feeder schools as a target market. The most common method to identify the target market is the Recency-Frequency-Monetary (RFM) model combined with data mining techniques. The original RFM model measures customer value by business profitability and focuses more on getting customers with high financial gains. However, educational institutions have different profit values. Therefore, this study adapts RFM model to the HEI marketing context to quantify feeder schools' value as a target market. Besides, geographical information is vital in marketing because knowing customers' geographical distribution enables businesses to devise more effective marketing strategies, resulting in more targeted marketing strategies. Therefore, this study includes geographic information. Besides, only a small amount of research, especially in the education field, has combined RFM model with geographic information and analyzed it spatially. Therefore, this study proposes a framework for determining feeder schools as a geomarketing target market using a modified RFM model combined with geographic information. This study modified RFM model into the Recency-Length-Frequency-CGPA-Persistence (RLFCP) model and combined it with the geographic information about the potential of districts (D) derived using spatial analysis and K-means clustering to obtain RLFCP-D model. Based on a case study at an Indonesian private university, this study chose two clusters with the highest values as the target segment. The target segment has three different enrollment patterns based on time series clustering. The decision tree model was used for profiling. The model's accuracy was 93.19% and 82.65%, respectively, based on validation and testing. This study recommends 141 feeder schools for prioritized target markets. These schools are 6.35% of the university feeder schools, contributing 54.95% of enrolled students. They are in 25 cities and 13 regencies from 19 provinces. The university should prioritize promotion to private schools while also targeting public schools in specific districts. This study contributes a data analytics framework that can help decision-makers choose feeder schools as the target market while paying attention to the location for developing a geomarketing strategy.

**MODEL RECENCY-FREQUENCY-MONETARY DIUBAHSUAI UNTUK
MENENTUKAN SEKOLAH SALURAN SEBAGAI SASARAN PROMOSI
GEOMARKETING DI INDONESIA**

ABSTRAK

Promosi ke pasaran sasaran yang tepat membantu institusi pengajian tinggi (IPT) meningkatkan kejayaan pengambilan pelajar dan menjimatkan kos. Sekolah menengah merupakan penyalur pelajar IPT. Oleh itu, melakukan promosi ke sekolah saluran adalah penting. Walau bagaimanapun, disebabkan oleh banyaknya bilangan dan meluasnya kawasan sekolah saluran, khususnya di Indonesia, pengurusan IPT perlu memilih sekolah saluran yang berimpak tinggi sebagai sasaran pasaran mereka. Kaedah yang paling biasa untuk mengenal pasti sasaran pasaran ialah model Recency-Frequency-Monetary (RFM) yang digabungkan dengan teknik perlombongan data. Asasnya, model RFM adalah untuk mengukur nilai pelanggan dengan keuntungan perniagaan dan lebih menumpukan kepada bagaimana mendapatkan pelanggan dengan keuntungan kewangan yang tinggi. Walau bagaimanapun, institusi pendidikan mempunyai nilai keuntungan yang berbeza. Oleh itu, kajian ini menyesuaikan model RFM kepada konteks pemasaran IPT untuk mengukur nilai sekolah saluran sebagai sasaran pasaran. Selain itu, maklumat geografi adalah penting dalam pemasaran kerana dengan mengetahui distribusi geografi pelanggan akan membolehkan perniagaan merangka strategi pemasaran yang lebih berkesan, dan lebih tersasar. Oleh itu, kajian ini merangkumi maklumat geografi. Walau bagaimanapun, hanya sejumlah kecil penyelidikan, terutamanya bidang pendidikan, yang telah menggabungkan model RFM dengan maklumat geografi dan menganalisisnya secara spatial. Oleh itu, kajian ini mencadangkan rangka kerja untuk menentukan sekolah saluran sebagai sasaran pasaran geomarketing dengan menggunakan model RFM yang diubahsuai serta digabungkan dengan maklumat geografi. Kajian ini mengubahsuai model RFM ke dalam model Recency-Length-Frequency-CGPA-Persistence (RLFPC) dan menggabungkannya dengan maklumat geografi: potensi daerah (D), hasil dari analisis spatial dan gugusan K-means, bagi memperoleh model RLFPC-D. Berdasarkan kajian kes di universiti swasta Indonesia, kajian ini memilih dua kelompok dengan nilai RLFPC-D tertinggi sebagai segmen sasaran. Segmen sasaran mempunyai tiga corak pendaftaran berbeza berdasarkan pengelompokan siri masa. Model pepohon keputusan digunakan untuk pemprofilan. Ketepatan model adalah 93.19% dan 82.65%, berdasarkan pengesahan dan ujian. Kajian ini mengesyorkan 141 sekolah saluran sebagai sasaran pasaran utama. Sekolah-sekolah ini adalah 6.35% daripada sekolah saluran universiti, yang menyumbang 54.95% daripada pelajar yang mendaftar. Mereka berada di 25 bandar dan 13 kabupaten dari 19 wilayah. Universiti harus mengutamakan promosi ke sekolah swasta di samping menyasarkan sekolah awam di daerah tertentu. Kajian ini menyumbang rangka kerja analisis data yang boleh membantu pembuat keputusan memilih sekolah saluran sebagai sasaran pasaran sambil memberi perhatian kepada lokasi untuk membangunkan strategi geomarketing.

ACKNOWLEDGEMENTS

I want to express my heartfelt gratitude to the following individuals and organizations for their support in finishing this thesis:

- Gs. Dr. Safiza Suhana Kamal Baharin, my main supervisor, for the supervision, suggestions, support, and trust so I can complete this thesis.
- Dr. Fauziah Kasmin, my co-supervisor, for the encouragement, supervision, suggestion, and motivation to complete this thesis.
- Universitas Atma Jaya Yogyakarta, for the financial and non-financial support to pursue my Ph.D.
- Universiti Teknikal Malaysia Melaka, for the opportunity and facilities in taking the Ph.D. program.
- Gs. Dr. Othman bin Mohd and Assoc. Prof. Gs. Dr. Asmala Ahmad, for all the thoughtful comments and suggestions in the proposal defense and progress presentations.
- Prof. Gs. Ts. Dr. Nanna Suryana Herman and Assoc. Prof. Ts. Dr. Ahmad Naim Che Pee, for accompanying us, UIPP batch 4, from the beginning of the program until we finish it.
- All my friends at Universitas Atma Jaya Yogyakarta, for the support.
- All my classmates, UIPP batch 4, for the togetherness, which is full of laughter.

TABLE OF CONTENTS

	PAGE
DECLARATION	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xii
LIST OF APPENDICES	xiii
LIST OF PUBLICATIONS	xiv
CHAPTER	
1. INTRODUCTION	1
1.1 Introduction	1
1.2 Research motivation	1
1.3 Research background and context	3
1.4 Problem statement	7
1.5 Research questions	9
1.6 Research objectives	9
1.7 Research hypotheses	11
1.8 Research scopes	11
1.9 Research contributions	11
1.10 Thesis outline	12
2. LITERATURE REVIEW	14
2.1 Introduction	14
2.2 Market segmentation	14
2.2.1 Demographic segmentation	18
2.2.2 Geographic segmentation	19
2.2.3 Psychographic segmentation	19
2.2.4 Behavioral segmentation	19
2.3 RFM-based model market segmentation	21
2.4 Data mining in market segmentation	25
2.4.1 Clustering	27
2.4.2 Classification	33
2.5 RFM-based market segmentation using data mining	37
2.6 RFM-based model application in the education domain	50
2.7 Geomarketing	57
2.7.1 Geographic information systems and spatial analysis	60
2.7.2 Geographic segmentation and spatial analysis	62
2.7.3 Spatial analysis in the education domain	65

2.8	Summary	67
3.	METHODOLOGY	68
3.1	Introduction	68
3.2	Research design	68
3.3	Research methodology	70
3.3.1	Preliminary study	74
3.3.2	Modify the RFM variables	74
3.3.3	Developing feeder school segmentation framework	75
3.3.4	Applying the proposed framework	75
3.3.5	Evaluation and recommendation	81
3.4	Limitation of the proposed methodology	81
3.5	Summary	81
4.	THE PROPOSED FRAMEWORK USING THE MODIFIED RFM MODEL	83
4.1	Introduction	83
4.2	The modified RFM model	84
4.2.1	RLFCP model	85
4.2.2	RLFCP scoring	87
4.3	The proposed framework	89
4.3.1	The first phase	91
4.3.2	The second phase	92
4.3.3	The third phase	92
4.3.4	The fourth phase	98
4.4	Summary	99
5.	EMPIRICAL RESULTS AND DISCUSSION	100
5.1	Introduction	100
5.2	Applying the proposed framework: a case study	100
5.2.1	Data acquisition	103
5.2.2	Data preprocessing and description	103
5.2.3	Feeder school segmentation	106
5.2.4	Validation and testing	146
5.3	Evaluation and recommendation	148
5.4	Discussion	149
5.4.1	Research question 1	149
5.4.2	Research question 2	150
5.4.3	Research question 3	154
5.4.4	Research question 4	156
5.4.5	Research question 5	163
5.4.6	Implication for practice	168
5.5	Summary	175
6.	CONCLUSION AND RECOMMENDATIONS	176
6.1	Conclusion	176
6.2	Recommendations	179

REFERENCES
APPENDICES

180
205



LIST OF TABLES

TABLE	TITLE	PAGE
1.1	Summary of the research problems	8
1.2	Relation between the research problems, questions, and objectives	10
2.1	Confusion matrix	36
2.2	The previous studies used RFM-based model and data mining techniques	38
2.3	The previous RFM-based model studies in the educational domain	53
2.4	The previous studies on geo-segmentation	63
3.1	Relation between research problem, research question, research objective, methodology and outcome	72
4.1	The RLFCP variables and definition	87
5.1	Data tables used in study	105
5.2	The number of enrolled students before and after data preprocessing	106
5.3	Distribution of feeder school regencies/cities in each province	108
5.4	Regencies/cities with the highest number of feeder schools	112
5.5	The number of feeder schools based on school ownership and type	113
5.6	The number of enrolled students based on school ownership and type	114
5.7	Regencies/cities with the highest number of enrolled students	117
5.8	The top 20 schools with the highest number of enrolled students	121
5.9	The statistics summary of the district's potential dataset	126

5.10	The Silhouette index for determining the optimal number of district's potential category	127
5.11	District's potential clusters statistics	128
5.12	Regency/city in the higher potential clusters	129
5.13	The RLFCP-D data for feeder school grouping	132
5.14	The RLFCP-D statistics	134
5.15	Silhouette index for determining the optimal number of feeder school clusters	135
5.16	The RLFCP-D characteristics of the feeder school clusters	137
5.17	The annual number of enrolled students at the target segment's feeder schools	140
5.18	The Silhouette index for determining the optimal number of enrollment patterns	140
5.19	The RLFCP-D statistics of the time series clusters	142
5.20	The validation results	147
5.21	The testing results	147
5.22	The number of samples based on the stratified sampling technique	148
5.23	The number of matched and mismatched cases of evaluation results	148
5.24	Comparison of the RLFCP model and its derivatives	154
5.25	Comparison of the RLFCP and RLFCP-D model	166
5.26	The recommended feeder schools for geomarketing promotion target	169

LIST OF FIGURES

FIGURE	TITLE	PAGE
2.1	The relationship triangle in student recruitment (Beneke and Human, 2010)	16
2.2	The STP process (DeSarbo et al., 2009)	18
3.1	Research methodology	71
3.2	Study area	79
4.1	The RLFCP model	86
4.2	The proposed framework	91
4.3	The flowchart of the segmentation phase	93
4.4	The RLFCP-D model	96
4.5	Prediction of feeder school segment	99
5.1	X University location	102
5.2	The number of data included for analysis	104
5.3	The feeder schools' distribution	107
5.4	The number of feeder schools from each province	109
5.5	The district classification based on the number of feeder schools	111
5.6	The Java's districts classification based on the number of feeder schools	112
5.7	The number of enrolled students in the analysis period	114
5.8	The number of enrolled students from each province	115
5.9	The district classification based on the number of enrolled students	116

5.10	The Java's districts classification based on the number of enrolled students	118
5.11	The classification of feeder schools based on the number of enrolled students	119
5.12	The classification of feeder schools on Java Island based on the number of enrolled students	120
5.13	The result of Global Moran's I test for the randomness of the number of feeder schools	122
5.14	Feeder school concentration	124
5.15	Feeder school concentration around D.K.I. Jakarta	125
5.16	Histogram of districts' number of feeder schools and enrolled students	126
5.17	The Elbow method for determining the optimal number of the district's potential category	127
5.18	District classification based on the district's potential category	130
5.19	District classification based on the district's potential category on Java Island	131
5.20	Histogram of the RLFCP-D variables	133
5.21	The Pearson's correlation coefficients between the RLFCP-D variables	134
5.22	Elbow method for determining optimal numbers of feeder school clusters	135
5.23	The result of the K-prototypes clustering	136
5.24	The result of ANOVA test	139

5.25	The result of the time series clustering	141
5.26	Time series enrollment pattern	142
5.27	The predictive decision tree model	143
5.28	The most prioritized feeder schools' distribution	145
5.29	The most prioritized feeder schools' distribution on Java Island	146
5.30	The priority school districts for the promotional target market	173
5.31	The priority school districts on Java Island for the promotional target market	174



LIST OF ABBREVIATIONS

ANOVA	-	Analysis of Variance
CART	-	Classification and Regression Tree
CGPA	-	Cumulative Grade Point Average
CLV	-	Customer Lifetime Value
CVI	-	Cluster Validity Index
DTW	-	Dynamic Time Warping
GPA	-	Grade Point Average
FN	-	False Negative
FP	-	False Positive
GIS	-	Geographic Information Systems
HEI	-	Higher Education Institution
PAM	-	Partitioning Around Medoids
RFM	-	Recency-Frequency-Monetary
RO	-	Research Objective
RP	-	Research Problem
RQ	-	Research Question
RFM	-	Recency-Frequency-Monetary
RLFCP	-	Recency-Length-Frequency-CGPA-Persistence
RLFCP-D	-	Recency-Length-Frequency-CGPA-Persistence-District's potential
St. dev	-	Standard deviation
TN	-	True Negative
TP	-	True Positive
WSS	-	Within-cluster Sum of Square errors

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	District's potential dataset and its category	205
B	The RLFCP-D dataset and its cluster	209
C	The annual number of enrolled students of the target segment's feeder schools	230



LIST OF PUBLICATIONS

1. Ernawati, Baharin, S.S.K., Kasmin, F. and Purwanugraha, H.A., 2022. Geo-marketing Promotional Target Selection using Modified RFM with Spatial and Temporal Analysis: A Case Study. *Journal of System and Management Sciences*, 12(3), pp. 156-180.
2. Ernawati, Baharin, S.S.K. and Kasmin, F., 2022. Modified RFM Model and the FUCOM Method for Identifying Higher Education Institution Target Market. *ICIC Express Letters*, 16(9), pp. 993-1000.
3. Ernawati, Baharin, S.S.K. and Kasmin, F., 2022. Target Market Determination for Information Distribution and Student Recruitment Using an Extended RFM Model with Spatial Analysis. *Journal of Distribution Science*, 20(6), pp. 1-10.
4. Ernawati, Baharin, S.S.K. and 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), pp. 87-112.
5. Ernawati, Baharin, S.S.K. and Kasmin, F., 2021. A Review of Data Mining Methods in RFM-based Customer Segmentation. *Journal of Physics: Conference Series 1869 012085*, pp. 1-8.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter serves as an introduction to this thesis. The motivation of this study is explained in the first section, followed by the context and background of the research. Following that, a declaration of the problem statements, objectives, questions, hypotheses, and scopes of this research are presented. After that, the study's contributions and structure organization are described.

1.2 Research motivation

Higher education institutions (HEIs) face increasing challenges (Kuiper, 2018; Miguéis et al., 2018). In Indonesia, with many HEIs (3,166 institutions under the Ministry of Research, Technology and Higher Education, 3,044 of them are private institutions (BPS Indonesia, 2021)), the challenge in student recruitment is intensifying (Jati et al., 2021). In this highly competitive environment, it is critical to develop and implement an effective marketing strategy (Pardiyono et al., 2022; Sukoroto et al., 2020), including promotion as a marketing element. One of the most heavily emphasized promotional activities that significantly disseminate university program information is direct promotion, namely a face-to-face activity with potential students, such as school visits and university open days (Chase et al., 2019; Messah and Immaculate, 2011). Visiting high schools is one way of promotion widely used by universities, especially private universities in Indonesia (Jati et al., 2021; Karna et al., 2015; Vernanda et al., 2019).

High schools are HEI stakeholders that serve as students' feeders (Seres et al., 2019) because their fresh graduates are the primary customers of the institution. Collaboration with high schools for marketing and student recruitment is common in Indonesia (Hidayat et al., 2020; Kartikasari, 2017). Building relationships with feeder schools is critical for HEI because these schools provide access to the institution's target market (Beneke and Human, 2010). Maintaining good communication with feeder schools keeps the institution in touch, so their trust and loyalty will increase.

Indonesia has 27,930 senior and vocational high schools across 34 provinces and 514 regencies/cities (BPS Indonesia, 2021), which could be potential markets for student recruitment promotion, particularly for most universities in Daerah Istimewa Yogyakarta (D.I. Yogyakarta) Province, where many out-of-town students study. However, reaching all high schools requires very large promotional resources because they are large in number and spread over wide areas. Therefore, university management must select which market segments to target and how to serve them (Barber et al., 2013).

Market segmentation is a popular approach that distinguishes customer groups based on customer features (Abbasimehr and Shabani, 2019). Market segmentation is used to identify the customers' characteristics, define the target customers, and provide information for developing marketing activities (Suhaibah et al., 2016). Traditionally, market segmentation analysis relies upon data from customer surveys (Kuiper, 2018). However, the adoption of information technology in many business processes generates a large amount of data that can be used for analytics in modern management to support optimal decision-making (Mekvabidze, 2020). Hence, companies can perform data analytics using data mining techniques to mine hidden knowledge from their customer's data, implement more efficient market segmentation strategies, and personalize promotional offers (Roshan and Afsharinezhad, 2017). Likewise, HEIs have a wealth of student transaction data for targeted

analytics (Daniel, 2015), such as enrollment and academic data. Accordingly, data mining can be employed to explore students' historical data to find targeted customers (Abaya and Gerardo, 2013). Ghosh et al. (2008) recommended using student registration data to segment the higher education market because it is relatively reliable and inexpensive. So the motivation of this study is to utilize HEI's transaction data for data analytics, which can assist HEI management in determining the target market and developing promotion strategies through market segmentation.

1.3 Research background and context

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 will receive the information is critical in promotional activities. Hence, determining the right target market becomes important to increase new customers' engagement and retention of current customers. The target segment consists of potentially valuable customers to whom concentrated promotional efforts may be applied (Roshan and Afsharinezhad, 2017). The market segmentation method can determine the target market, which provides significant success in the higher education market (Lewison and Hawes, 2007).

Market segmentation separates customers into smaller, homogeneous, distinct, and specific customer groups based on customer features (Abbasimehr and Shabani, 2019; Tsiptsis and Chorianopoulos, 2009). Studies on market segmentation have been widely applied in various fields, for example, e-commerce (Beheshtian-Ardakani et al., 2018; Dachyar et al., 2019; Kadir and Achyar, 2019), medical (Hosseini and Mohammadzadeh, 2016; Tarokh and EsmaeiliGookeh, 2019), financial (Abbasimehr and Shabani, 2019; Aryuni et al., 2018; Firdaus and Utama, 2021; Sheikh et al., 2019), and education (Abaya

and Gerardo, 2013; Davari et al., 2018; Hidayat et al., 2020; Kartikasari, 2017; Purfini and Yunanto, 2019; Wong and Chong, 2018). The recent customer segmentation study has frequently adopted the Recency-Frequency-Monetary (RFM) model for customer segmentation and targeting (Christy et al., 2021; Firdaus and Utama, 2021; Hwang and Lee, 2021; Kit et al., 2021; Wu et al., 2020). RFM is a powerful method in marketing (Gustriansyah et al., 2020) that has a high ability to find potential customers (Hwang and Lee, 2021). This model can identify valuable customers based on purchase behavior (Wei et al., 2010) and reveal customers' categories targeted for personalized services in marketing campaigns (Vitor et al., 2018). RFM has three variables: recency, frequency, and monetary. Recency expresses the novelty of customer relations with the company. Frequency is related to the number of times customers transacted, whereas monetary value represents the number of money customers paid during the analysis period. A customer who has recently and frequently transacted with large sums of money is considered more valuable and profitable to become a target market.

RFM model captures customer characteristics with fewer segmentation variables (Chen et al., 2012; Rezaeinia and Rahmani, 2016), making it simple, easy to implement, and easy to understand by managers and decision-makers (Peker et al., 2017; Sarvari et al., 2016). Furthermore, the values of each RFM factor are easy to get from transactional databases, and the model works well with transactional data, making the results easy to interpret (Carneiro and Miguéis, 2021). However, RFM model is incomplete (Moghaddam et al., 2017), and simplicity weakens its power (Tavakoli et al., 2018). Therefore many studies suggest modifying or extending the model by adding other useful and critical variables (Dedi et al., 2019; Hidayat et al., 2020; Moghaddam et al., 2017; Peker et al., 2017; Singh and Singh, 2016; Tarokh and EsmaeiliGookeh, 2019; Wei et al., 2019). Besides, RFM model is not always proper for all application domains. Hence, businesses may create a