Identifying Market Segment of University Student using Cluster Analysis Based on Psychographics Variable

Hardika Khusnuliawati
Universitas Sahid Surakarta
Surakarta, Indonesia
khusnuliawati@usahidsolo.ac.id

Dhian Riskiana Putri
Universitas Sahid Surakarta
Surakarta, Indonesia
dhianrp@gmail.com

***Abstract*—The importance of the presence of higher education enables the private sector to participate in organizing academic activities in the form of higher education institutions. This causes the higher education market to become more competitive, which implies a low number of students. Therefore, market segmentation needs to be applied to college students so that they can help to determine the model of marketing and promotional activities. The stages carried out in this study consisted of data collection, data exploration, and extracting segment. Cluster analysis was applied as a method for extracting segments of students with psychographics variables as partitioning factors. The K-Means algorithm was chosen as the method applied for cluster analysis because it produces better performance when compared to the use of K-Modes. Cluster analysis based on psychographics variables applied to this case succeeded in extracting the segment of the university students into 6 segments.**

Keywords—market segmentation, cluster analysis, psychographics variable

I. INTRODUCTION

The presence of higher education hold one of the most important roles in shaping the future of society. Apart from public Higher Education Institutions (HEI), many countries provide opportunities for the provision of HEI from the private sector. The presence of private HEI are expected to expand people's opportunities for obtaining higher education. However, the increasing number of private HEI has also contributed to increasing their competition for students [1]. Therefore, the use of marketing strategies is necessary for HEI.

There are three main activities in the marketing strategy called the holy trinity of marketing [2]. These activities consist of segmentation, targeting and positioning (STP). Of the three stages, market segmentation is the first step in creating a buying leading strategy which determines the results of the following stages [3]. Market segmentation has been widely applied in marketing and promotion by business companies to better understand what customers need and want. In [4], market segmentation is used in customer relationship management (CRM) for multimedia on demand. The use of market segmentation is also applied to financial services targeting saving customers [5]. In addition, manufacturing and trade enterprises operating in the coffee market uses market segmentation to identify needs and expectations of consumers [6].

Whereas in educational services, market segmentation can be used as a guide for school or college management to build an effective communication with prospective students. HEI can use market segmentation to better understand the needs and expectations of prospective students so that the services provided by higher education are more valuable to them [7]. Market segmentation can also be used to support promotional strategies in attracting prospective school students [8]. In addition, from the results of market segmentation, hey can build an effective form of communication through various media based on the characteristics of each market segment [9].

Determining the variables to be used is part of market segmentation that needs attention. In the case of implementing market segmentation for HEI, the selected variables are those that have the potential to uncover the characteristics of each market segment with their measurability, availability, reliability, and ability [10]. One of the variables that can be used is the psychographic variable. This variable significantly improves the explanatory power of e-consumer preferences [11]. Market segmentation based on psychographic variables could also be a powerful way to reach people from otherwise different demographics groups, geographies or other characteristics by highlighting the specific product properties that are in interest of the particular segment [12]. For application in HEI, psychographic variables have a big influence on the results of the segmentation so that they will be used for this study.

. In addition to determining the variables, the selection method of segmentation is another aspect to consider. The method that has been commonly used in market segmentation is cluster analysis [13], [14]. K-Means is one of the cluster analysis methods applied to market segmentation for HEI. It is able to give the most meaningful results to describe prestige-seeking university students [15].

This paper aims to identifying market segment of university student using cluster analysis based on psychographic variable. Market segmentation applied to HEI aims to produce profiles of targeted students based on psychological factors that help to determine the model of marketing and promotional activities more accurate and efficient. K-Means will be used in the study as a cluster analysis method implemented in the HEI market segmentation. Another clustering algorithm, K-Modes, will be compared with k-means to see which cluster analysis method gives the best results.

II. LITERATURE REVIEW

This section will discuss methods related to the research conducted.

A. Market Segmentation

Market segmentation is one of the keys in building strategic marketing which can be a decision-making tool for marketing managers [16]. In [17], market segmentation can be defined as the actual process of identifying segments of the market and the process of dividing a broad customer base into sub-groups of consumers consisting of existing and

prospective customers. Market segmentation helps to understand what consumers need and want so that the marketed products can satisfy consumers.

The sequence of segmentation method follows: collect information or data relating to the customers, determine the differences and similarities between customers, and finally, divide the customer base into groups with similar behavior [18]. Market segmentation can be applied by different practitioners in different domains and produce satisfactory results.

B. Psychographic Variable

Psychographic variables can be used in market segmentation according to the personality traits, values, motives, interests and lifestyle of consumers. Psychographic variables are often used when buying behavior correlates with a consumer's personality or lifestyle [19].

Angulo (2015) represents psychographic variables as emotional perspectives such as personal value for market segmentation in HEI [7]. Whereas, Goodrich (2018) uses a list of questions that represent satisfaction with academic, college, and personal life, as well as in different characteristics of students desired at a school during their college planning stage as a psychographic variable for market segmentation [9]. The use of psychographic variables also has a role in finding the major factors students emphasized when selecting a school or department [20]. Usually, to obtain psychographic variables, a questionnaire is used which contains a list of questions.

C. Cluster Analysis

Cluster analysis is the field used for extract market segments as a segmentation method [16]. Consumers will be sorted into relatively homogenous groups according to chosen variable, so that consumers placed together in the same cluster, show more similarities with each other than with those placed in other clusters [14]. Several clustering methods that will be applied in this research include:

1) K-Means.

K-means clustering is one of the non-hierarchical data clustering methods that has several advantages, such as simple and fast [21]. The steps of clustering with K-Means method are as follows [8]:

a) Select the number of clusters k and initialize the cluster center k. Cluster centers are assigned initial values with random numbers.

b) Allocate all data objects to the nearest cluster. The proximity of a data to a particular cluster is determined the distance D_{ij} between the data X_{ki} with the cluster center X_{kj}. To distance all data to each cluster center point can use Euclidean distance theory formulated as follows equation (1):

$$D_{ij} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad (1)$$

c) Recalculate cluster center with current cluster membership. The new cluster center is the average of all data in a particular cluster.

d) Reassign each object using the new cluster center. If the cluster center does not change again then the clustering process is complete.

2) K-Modes

K-Modes clustering algorithm is an extension to the standard K-Means clustering algorithm that operates on categorical data [22]. The modifications done in the K-Means includes using a simple matching dissimilarity measure for categorical objects, replacing means of clusters by modes, and using a frequency-based method to update the modes [23]. The dissimilarity measure can be defined using equation (2).

$$(X, Y) = \sum_{j=1}^{m} \delta(x_j, y_j) \quad (2)$$

where,

$$\delta(x_j, y_j) = \begin{cases} 1, & x_j \neq y_j \\ 0, & x_j = y_j \end{cases} \quad (3)$$

For X, Y are two categorical objects described by m categorical attributes.

The evaluation method of cluster results uses the Davies-Bouldin Index (DBI). It represents how well clustering has been done by calculating the quantity and the derivative features of the dataset [24]. The better cluster results have a value close to non-negative zero.

III. Implementation and Analysis

In the following section, the stages of identifying market segments of university students using a cluster analysis based on psychographic variables are described as follows.

A. Data Collecting and Preprocessing

Data were collected from 230 students from Sahid University of Surakarta as one of the private university in the region of Central Java, Indonesia. The resulting dataset from the data collection consists of 230 objects with a total of 36 features. These features are the answers to questionnaires from students that represent psychographic variables. Each questionnaire item represents one feature. All items in the questionnaire were arranged on 5-point Likert-type scales (1 = completely disagree to 5 = completely agree).

The list of questions on the questionnaire is divided into 7 categories which are described in the Table 1. Every category is represented by a number of items in the form of questions that are listed in the distributed questionnaire.

After the data is collected, preprocessing is applied to the dataset. Preprocessing includes data cleaning and data transformation [25]. For data cleaning, At the data cleaning stage, the data rows that have missing values are filled with the average value of the attributes. Then, data features with categorical types are transformed into numeric value. Table 2. shows an example of a dataset used for the cluster analysis stage. The row shows the id of the student while the column shows the question items displayed on the questionnaire.

B. System Design and Implementation

In this research, there are several steps that need to be done, as shown in the Fig. 1. The first stage is collecting data from higher education students with the selected variable. Survey data using questionnaire were collected to obtain psychographic variable.

TABLE I. A LIST OF CATEGORIES THAT REPRESENT THE PSYCHOGRAPHIC VARIABLES OF THE QUESTIONNAIRE

No.	Categories	Code	Description
1.	Orientation for Work	A	Desire to easily get a job and expand the network for work.
2.	Quality from the university	B	The quality of the university such as national and international rankings, the awards they have, the cooperation they have, the name of a famous campus..
3.	Personal value	C	Attracted by the atmosphere of the campus.
4.	Motivation from family	D	Encouragement from family to go to college.
5.	Economy	E	Affordable tuition fees and the availability of scholarships.
6.	Effect of Advertising and promotions	F	Results from campus promotions and recommendations from others.
7.	Supporting external factors	G	Strategic campus location and supporting living facilities around the campus environment.

TABLE II. THE SAMPLE DATASET FROM PREPROCESSING RESULTS

ID	A1	A2	A3	…	G3	G4	G5
1	5	4	5	…	5	5	5
2	5	4	4	…	4	3	4
3	5	4	5	…	5	4	2
4	5	2	4	…	4	4	3
5	5	4	4	…	5	2	4
	…						
230	5	4	5	…	4	3	5

Preprocessing is performed at the data exploration stage. The results of the preprocessing stage are the dataset without missing values and has been transformed into numeric values.

At the extracting segments stage, cluster analysis is applied to the data obtained from the previous stage. The K-Means and K-Modes algorithms are used as methods for performing cluster analysis. The two algorithms are evaluated and the results are compared. Based on the best results from the cluster analysis, the student segment profile is determined according to the label of the clustering results. Then, examine the characteristics of each student segment.

C. Results and Analysis

The evaluation of market segmentation with cluster analysis is done by finding the DBI value and length of time processing. The results from Table 1. show a small value of DBI, so it can be concluded that the cluster analysis for market segmentation of university student performs a good clustering. The performance of K-Means algorithm also exceeds the K-Modes algorithm in term of the DBI value and time processing. K-means produces a smaller DBI with a value of 2.0604 compared to K-Modes with a value of 3.4786. In terms of processing time, K-Means takes less time. It takes 7 seconds to produce 6 clusters. Meanwhile, K-Modes takes 8 seconds with a result of 3 clusters.

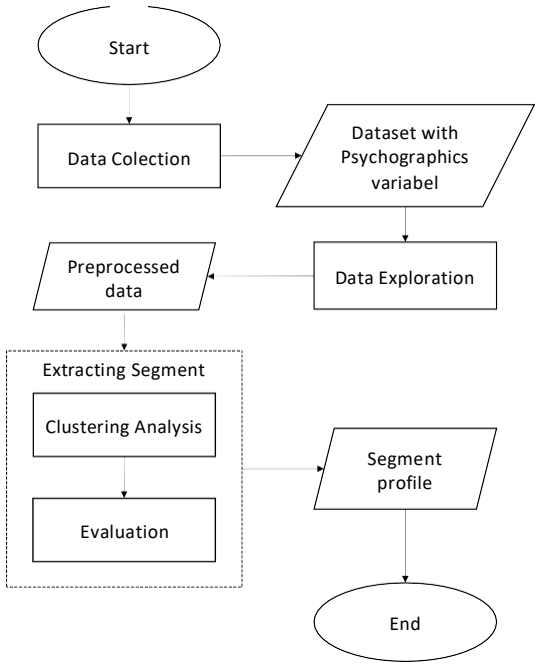

Fig. 1. Flowchart of clustering analysis based psychographics variable.

TABLE III. THE PERFORMANCE COMPARISON OF CLUSTERING ANALYSIS METHODS

Clustering methods	DBI	Time processing (second)	Number of cluster
K-Means	2.0604	0.0628	6
K-Modes	3.4786	0.1067	3

Based on the evaluation, the segment profile is examined using the clustering results of K-Means algorithm. Students are divided into six segments with the following characteristics.

- **Highly motivated due to economic factors (HE)**. Students in this cluster consider economic factors as the highest reason when choosing a university. Tuition fees and scholarship offers are the highest concern when compared to other factors.
- **Highly motivated because of family encouragement (HF)**. Family motivation is the highest factor for students in this cluster to continue their education to higher education. Other factors are less of a concern for them.
- **Motivated by the quality of universities but promotion does not reach them (MQ)**. Students in this cluster pay more attention to university quality factors when choosing higher education to continue their studies. However, compared to other clusters, students were less accessible by university promotions.
- **Motivated to get decent work by paying attention to personal values (MW)**. The highest factor in continuing their studies in higher education for

students in this cluster is getting a decent job by paying attention to their personal values.

- **Seekers of maximum benefit while in college (SB)**. Students in this cluster pay high attention to all psychographic factors when choosing higher education to continue their studies.
- **Neutral student (NS).** Students in this cluster gave medium scores in determining their motivation to choose higher education institutions based on psychographic factors.

Based on the segments obtained, it can be seen that the six segments have different characteristics. Fig 2. shows the characteristics of each segment based on the categories shown in Table 1. The segment profiles can be used as a reference for effective promotion of targeted student candidates.

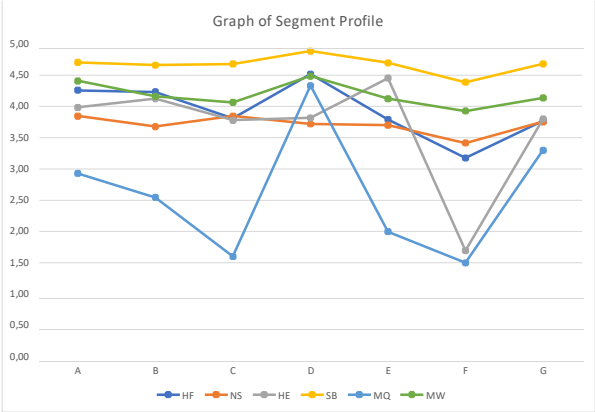

Fig. 2. Graph of segment profile of university students.

IV. Conclusion

The cluster analysis implemented in university students is able to produce segment profiles based on psychographic variables. The K-Means algorithm shows better c results than using K-Modes. This is indicated by a smaller DBI value and a shorter time processing.

The number of segments resulting from K-Means clustering is six. The characteristics of each segment include highly motivated because of family encouragement, highly motivated because of family encouragement, motivated by the quality of universities but promotion does not reach them, seekers of maximum benefit while in college, and neutral student**.**

For future work, the use of variables other than psychographics is required. So that, the information extracted from each segment can be richer. Moreover, another cluster analysis method can be proposed to find a new way to improve cluster results.

Acknowledgment

Authors thank Ministry of Education and Culture Republic of Indonesia and Universitas Sahid Surakarta.for supporting us regarding financial and data for this research.

References

[1] L. Leonnard, H. K. Daryanto, D. Sukandar and E. Z. Yusuf, "The loyalty model of private university student," *International Research Journal of Business Studies*, vol. 7, no. 1, 2015.

[2] B. B. Schlegelmilch, "Segmenting targeting and positioning in global markets," *Global marketing strategy*, pp. 63-82, 2016.

[3] M. Arsova and R. Temjanovski, "Strategy for market segmentation and differentiation: contemporary marketing practice," *Journal of Economics*, vol. 4, no. 1, pp. 27-35, 2019.

[4] C. Hung and C.-F. Tsai, "Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand," *Expert systems with applications*, vol. 34, no. 1, pp. 780-787, 2008.

[5] R. Makgosa, T. Matenge and P. Mburu, "Hybrid segmentation in the financial services market: targeting saving consumers.," *Family and Consumer Sciences Research Journal* , vol. 44, no. 4, pp. 447-468, 2016.

[6] G. Maciejewski, S. Mokrysz and Ł. Wróblewski, "Segmentation of coffee consumers using sustainable values: Cluster analysis on the polish coffee market," *Sustainability*, vol. 11, no. 3, p. 613, 2019.

[7] F. Angulo, A. Pergelova and J. Rial, "A market segmentation approach for higher education based on rational and emotional factors," *Journal of Marketing for Higher Education*, vol. 20, no. 1, pp. 1-17, 2010.

[8] I. Sari, A. Maseleno, F. Satria and M. Muslihudin, "Application model of k-means clustering: Insights into promotion strategy of vocational high school.," *International Journal of Engineering & Technology*, vol. 7, no. 2.27, pp. 182-187, 2018.

[9] K. Goodrich, K. Swani and J. Munch, "How to connect with your best student prospects: Saying the right things, to the right students, in the right media.," *Journal of Marketing Communications* , pp. 1-20, 2018.

[10] C. Lin, "Segmenting customer brand preference: demographic or psychographic," *Journal of Product & Brand Management* , 2002.

[11] H. Liu, Y. Huang, Z. Wang, K. Liu, X. Hu and W. Wang, "Personality or Value: A Comparative Study of Psychographic Segmentation Based on an Online Review Enhanced Recommender System," *Applied Sciences 9*, vol. 9, no. 10, p. 1992, 2019.

[12] M. Nadanyiova, L. Gajanova and J. Majerova, "THE IMPACT OF PSYCHOGRAPHIC SEGMENTATION ON THE PERCEPTION OF GREEN MARKETING," in *Economic and Social Development: Book of Proceedings*, 2020.

[13] R. Govindasamy, S. Arumugam, J. Zhuang, K. M. Kelley and I. Vellangany, "Cluster Analysis of Wine Market Segmentation-A Consumer Based Study in the Mid-Atlantic USA," *Economic Affairs*, vol. 63, no. 1, pp. 151-157, 2018.

[14] H. Müller and U. Hamm, "Stability of market segmentation with cluster analysis–A methodological approach," *Food Quality and Preference* , vol. 34, pp. 70-78, 2004.

[15] R. Casidy and W. Wymer, "A taxonomy of prestige-seeking university students: strategic insights for higher education," *Journal of strategic marketing*, vol. 26, no. 2, pp. 140-155, 2018.

[16] S. Dolnicar, B. Grün and F. Leisch, "Market Segmentation Analysis.," *Market Segmentation Analysis*, pp. 11-22, 2018.

[17] M. A. Camilleri, "Market segmentation, targeting and positioning," *Travel Marketing, tourism economics and the airline product*, pp. 69-83, 2018.

[18] P. W. Murray, B. Agard and M. A. Barajas, "Market segmentation through data mining: A method to extract behaviors from a noisy data set," *Computers & Industrial Engineering*, vol. 109, pp. 233-252, 2017.

[19] G. Armstrong, S. Adam, S. Denize and P. Kotler, Principles of marketing., Pearson Australia, 2014.

[20] Y.-F. Chen and C.-H. Hsiao, "Applying market segmentation theory to student behavior in selecting a school or department," *New Horizons in Education*, vol. 57, no. 2, pp. 32-43, 2009.

[21] C. Ordonez and E. Omiecinski, "Efficient disk-based K-means clustering for relational databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 8, pp. 909-921, 2004.

[22] D. Kamthania, A. Pawa and S. S. Madhavan, "Market Segmentation Analysis and Visualization Using K-Mode Clustering Algorithm for

E-Commerce Business.," *Journal of computing and information technology*, vol. 26, no. 1, pp. 57-68, 2018.

[23] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283-304, 1998.

[24] F. Tempola and A. F. Assagaf, "Clustering of Potency of Shrimp in Indonesia with k-means algorithm and validation of davies-bouldin index," in *International Conference on Science and Technology (ICST 2018)*, 2018.

[25] P. D. Hung, N. D. Ngoc and T. D. Hanh, "K-means Clustering Using RA Case Study of Market Segmentation," *Proceedings of the 2019 5th International Conference on E-Business and Applications*, pp. 100-104, 2019.