Jurnal CoreIT, Vol.9, No.1, June 2023
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Segmentation of Mentoring Customer Characteristics Using the K-Means Method and Hierarchical Clustering for Customer Relationship Management (CRM)

Hanif Aristyo Rahadiyan[1*], Mujib Ridwan[2], Anang Kunaefi[3]
[1,2,3] Dept. of Information System, Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia
h96218058@uinsby.ac.id (*Corresponding Author)

Abstract. In the next 10-20 years, it is expected that Indonesia will enter a demographic bonus era, where the population of productive age exceeds that of non-productive age. This presents an opportunity for startups in the field of education to prepare better human resources in Indonesia. With the recent Covid-19 pandemic, the government has implemented regulations that require online teaching and learning. Startups, such as Outstanding Youth Indonesia (OYI), play a role in bridging distance learning, leading to increased competition in the education sector. To stay competitive, OYI is implementing a customer relationship management (CRM) strategy, using consumer characteristic segmentation through the K-means method and hierarchical clustering. The study aims to test the consumer characteristic cluster results and provide CRM recommendations based on the segmentation results. The results of the study revealed that the K-Means method was more optimal, with a score of 0.657, compared to hierarchical clustering of 0.644. The clusters tested included categories, intended education, and types of scholarships. Three clusters were produced: cluster 1, dominated by high school/vocational high school students; cluster 2, mostly university students; and cluster 3, dominated by employees of government agencies. Cluster one had the largest silhouette coefficient. Based on the clustering, a strategy was generated for each cluster to improve CRM in OYI.

Keywords: CRM, Hierarchical Clustering, K-Means

Received January 2023 / **Revised** June 2023 / **Accepted** July 2023

INTRODUCTION

Indonesia is expected to enter a demographic bonus in the next 10-20 years, where the productive age population exceeds the population aged 65 years and over or non-productive age, which covers more than 64% of Indonesia's total population [1]. However, some problems may occur if the productive age population cannot compete with the market. If the human resources, especially those of productive age, are of good quality, it will undoubtedly help to boost economic growth and productivity in a country. However, if these resources are not properly prepared, it could have a negative impact [2]. Seeing this phenomenon, an opportunity has emerged that has the potential to become a business area, namely by establishing a startup in the education sector. One of the startups engaged in the development of education and youth empowerment is Outstanding Youth Indonesia (OYI). OYI offers self-development programs according to the needs of today's youth such as sharing, mentoring, and youth events so that they can increase the value of Indonesian youth.

The existence of a pandemic in recent years has changed people's habitual patterns in carrying out daily activities where online activities have become one of the mainstays in socializing remotely so that youth are given the convenience of access anywhere and anytime. This condition causes OYI to compete in providing different services to each customer depending on the nature and needs of customers based on the company's customer data [3]. It aims to improve service to consumers by implementing consumer segmentation which is one of the methods of the CRM phase to identify customer characteristics.

Segmentation activities aim to identify customer characteristics and implement appropriate marketing strategies for the company's benefit [4]. By using customer segmentation techniques, OYI can target customers' specific needs and identify the most profitable segments. This approach can improve customer loyalty by developing, offering, and providing tailored products and services, thus satisfying current customers and attracting new ones [5].

However, OYI faces several challenges, including inadequate processing of consumer data to extract valuable information for the company. The software currently in use is limited to data storage and lacks visualization tools to provide insights into business conditions, making it difficult for the company to develop strategic policies. To address these issues, a study was conducted to analyze OYI consumers, with the goal of identifying consumer segments through clustering techniques.

Clustering is an unsupervised learning technique in data mining that can be used to identify patterns in data sets [6]. There are two main approaches in clustering, hierarchical and non-hierarchical. Popular methods in hierarchical clustering include Single Linkage, Complete Linkage, Average Linkage, and Centroid Linkage. In non-hierarchical clustering, popular methods include K-means and Fuzzy C-means. In this research, the K-means method was chosen in non-hierarchical approach due to its superior performance over Fuzzy C-means in previous studies. In the Hierarchical approach, the Average Linkage method was chosen due to its higher silhouette score compared to other hierarchical methods in previous research.

After clustering, the quality and strength of the clusters are evaluated using the silhouette coefficient method. The silhouette coefficient, or silhouette score, is a technique used to evaluate how well an object fits within a cluster [7]. Lastly, an analysis based on the results of consumer segmentation is conducted using a Customer Relationship Management approach to enhance customer relationships. The segmentation results can be utilized by the marketing department in selecting promotional strategies that are tailored to the characteristics of potential customers, increasing the effectiveness and efficiency of promotions.

METHODS

1. **Data retrieval**
   At this stage, customer data from OYI is collected in a spreadsheet format for further analysis. The parameters used for this analysis include the registration date, name, telephone number, institution, intended level, type of scholarship, and proof of transfer.
2. **Cleaning Data**
   At this stage, data cleaning is conducted to prevent issues during data processing. The cleaning process is divided into three parts:
   - **2.1. Data Selection**
     The data selection process aims to prevent problems during data processing. Data selection is done by removing unnecessary attributes and attributes that contain private information, such as telephone numbers and emails.
   - **2.2. Pre-processing**
     At this stage, data is checked and repaired for issues such as null values, inconsistencies, redundancies, and other errors that can occur during data processing. Additionally, an enrichment process is conducted, which involves adding relevant external information.
   - **2.3. Transformation**
     The format of the data is modified to ensure smooth processing, for example, using the label encoder which assigns labels from strings to integers.
3. **Implementation of K-Means and Hierarchical clustering**
   The following are the stages of comparing K-Means and Hierarchical clustering to determine the optimal method.
   - **3.1. K-Means**
     K-Means is a clustering algorithm that divides data into groups or clusters, grouping similar data into one cluster and dissimilar data into separate clusters, based on their characteristics [8]. The first step in the K-Means algorithm is to determine the value of K, which represents the number of clusters. The centroids of these clusters can be determined randomly. Then, each point or object is assigned to the cluster whose centroid it is closest to. The following are the steps of the K-Means algorithm [9]:
     a. The number of clusters is determined by distributing the clusters randomly, then the optimal number of clusters is determined by using the Elbow method.
     b. The initial centroid of a cluster is randomly determined with a predetermined number of clusters.
     c. The distance between each data point and the centroid is calculated, with the data point being assigned to the cluster of the closest centroid. The distance is calculated using the Euclidean Distance equation [10].
     d. Each data point is grouped according to the nearest centroid.

e. The centroid value is then updated based on the average value of the cluster in question using the following formula.

$$C_k = \frac{1}{n_k} \sum d_i \qquad (1)$$

Explanation:

n_k = The number of data points in the cluster.

d_i = The sum of the distance values for all data points in the cluster.

f. If there are no changes in the membership of each cluster, the loop is considered completed. At this point, the average of the centroid is used as a parameter to determine the distribution of data.
g. If not, it will be repeated starting from point (b).

3.2. Hierarchical clustering

Hierarchical clustering is a method of grouping objects based on their similarity values, where objects with the closest similarity are grouped together [11]. The Hierarchical Clustering method is a technique for creating a hierarchy of data sets based on the similarity of object characteristics [12]. Here are some techniques used in the Hierarchical Clustering method:

a. Determine the method for calculating the distance of each cluster, in this study the Average linkage method is used to measure cluster proximity.
b. Each data point is considered a cluster, then the distance between each point is calculated using the Euclidean distance method.
c. Clusters with the smallest distance are merged into one cluster.
d. If the total number of clusters is equal to 1, the process is considered completed.
e. If the total number of clusters is not equal to 1, the distance between each new cluster is calculated using the Average linkage method with the following equation:

$$d_{(AB)C} = \frac{\sum u \sum vd_{uv}}{N_{(AB)} N_C} \qquad (2)$$

4. Testing the clusters

The Silhouette Coefficient method is used to determine the correctness of the obtained clusters. The procedure for evaluating the silhouette coefficient value is as follows:

a. Calculating the average distance from an object i to all objects within the same cluster.
b. The average value is taken and is referred to as ai.
c. For each object i, he average distance from object i to objects in other clusters is calculated.
d. The minimum value is taken from all resulting average distances, referred to as bi.
e. After obtaining the values of ai and bi, the silhouette coefficient is calculated using the following equation:

$$Si = \frac{(bi - ai)}{\max(bi, ai)} \qquad (3)$$

f. The Silhouette Coefficient score ranges from -1 to 1. A value closer to 1 indicates better quality of data grouping within a cluster..

5. Analyze and Evaluate the results

The final stage is to conduct the analysis and evaluate the results obtained from implementing clustering using both the K-means and Hierarchical clustering methods. A pattern or character analysis of OYI customers is performed by summarizing the attributes of each cluster.

RESULT AND DISCUSSION

These results describe of process outline the steps taken from data collection, data cleaning, segmentation using the K-means method, Hierarchical clustering, cluster results testing, and analysis of patterns or characteristics of consumers. Additionally, an evaluation was performed using the Customer Relationship Management (CRM) approach to enhance customer attractiveness and customer loyalty at Outstanding Youth Indonesia (OYI).

1. Data retrieval

Data retrieval is conducted directly from the company's Cloud Storage after consulting with the Chief Executive Officer (CEO) and Chief Technology Officer (CTO). The data obtained is in the form of excel files from each batch, with a total of 373 data. All data will then be cleaned to remove any data that does not meet the requirements. The OYI consumer data obtained includes the attributes of Date, Name, Telephone Number, Email, Institution, Intended Level, Type of Scholarship, and Proof of Payment. A sample of the OYI consumer data is in the figure below:

Tanggal	Nama	Nomor Telepon	Email	Institusi	Jenjang yang Dituju	Jenis Beasiswa	Bukti Transfer
4/20/2020	Mohammad fikri ramadhani fauzi	0852xxxx	daxx@xx	MA Unggulan Hikmatul Amanah	S1	Dalam Negeri	https://drive.google.com/xxxxx
4/21/2020	NUR ROKHIM	0856xxxx	nuxx@xx	IAIN TA	S2	Luar Negeri	https://drive.google.com/xxxxx
4/22/2020	Retno Puji Utami	0897xxxx	rexx@xx	SMK Negeri 1 Gombong	S1	Dalam Negeri	https://drive.google.com/xxxxx
4/22/2020	Nur rohmatul layly	0877xxxx	nuxx@xx	MAS Hikmatul amanah	S1	Dalam Negeri	https://drive.google.com/xxxxx
4/22/2020	Moh. Ulul Azmi	0856xxxx	azxx@xx	Universitas Indonesia	S2	Dalam Negeri	https://drive.google.com/xxxxx
4/22/2020	Alissa Nuriyah	0857xxxx	alxx@xx	SMA Darul Ulum 2 Jombang	S1	Dalam Negeri	https://drive.google.com/xxxxx
4/22/2020	Nor khofifah	0812xxxx	noxx@xx	SMA DU 1 unggulan BPPT jombang	S1	Dalam Negeri	https://drive.google.com/xxxxx

Figure 1. OYI Consumer Data Overview

2. Cleaning Data

At this stage, the entire data is checked and repaired or some data that is null, inconsistent, redundant, or other errors that can cause errors during the clustering process are deleted. This stage is divided into three parts:

2.1. Data Selection

OYI consumer data that will be processed consists of eight attributes, namely Date, Name, Telephone Number, Email, Institution, Target Level, and Proof of Payment. In the data selection process, attributes that are not necessary are identified and eliminated to facilitate processing. Several attributes are omitted, including Date, Name, Telephone Number, Email and Proof of Payment. As a result, from the data selection process, the required attributes are obtained: Institution, intended level and type of scholarship.

2.2. Pre-processing

The pre-processing stage is carried out to convert the raw data into a more efficient format so as to avoid obstacles at the data processing stage. The pre-processing consists of deleting some Null-valued data, etc.

2.3. Transformation

The data transformation stage uses the one-hot encoding method and label encoding with the scikit-learn library in Python. The data for each attribute is initialized with numbers starting from 0,1,2,3 and so on.

The data acquisition after going through the data cleaning stages is shown in Table 1 below:

Table 1. Result Cleaning Data

Index	Institution	Types of Scholarships	Educational Stage
Total	280	280	280
Unique	138	2	3
Modus	Universitas Brawijaya	Dalam Negeri	S2

After going through the data cleaning stages, the previous 373 consumer data was reduced to 280 after filtering out redundant and invalid data.

3. Implementation of K-Means and Hierarchical clustering

After the OYI consumer data has gone through the transformation stage and become numeric, it is processed using Python with the K-Means method and Hierarchical clustering with a Euclidean distance matrix. The clustering implementation consists of three stages: determining the optimal number of clusters using the Elbow method, grouping the data, and finally validating it using the Silhouette Coefficient method.

3.1. K-Means

a. Determine the optimal number of clusters

The Elbow method is a technique that can be used to determine the optimal number of clusters using the python programming algorithm. In Table 2 below, the SSE (Within-Cluster-Sum-of-Squares) of each experimental cluster is presented:

Table 2. SSE K-Means

Number of Clusters	SSE
2	69864,10
3	30302,37
4	18993,96
5	10532,40
6	7840,18
7	5970,62
8	4629,06
9	3476,64

After obtaining the SSE value for each cluster, a comparison of the differences between each cluster is carried out. The cluster with the most significant difference is considered the optimal cluster. The decline rate of each K (cluster) is presented in Table 3 below:

Table 3. The Rate of decline of SSE K-means

Kn - Kn+1	The rate of decline
K2 – K3	**39561,73**
K3 - K4	11308,41
K4 – K5	8461,56
K5 – K6	2692,22
K6 – K7	1869,56
K7 – K8	1341,56
K8 - K9	1152,42

In conclusion, the optimal number of clusters according to the Elbow method is 3 clusters. The graph of SSE value in the Elbow method is presented in Figure 2 below, where the X-axis represents the number of experimental clusters, ranging from 2 to 9 clusters, and the Y-axis represents the SSE value obtained from each K value from the X-axis. The Elbow graph for determining the optimal number of clusters is presented in Figure 2:

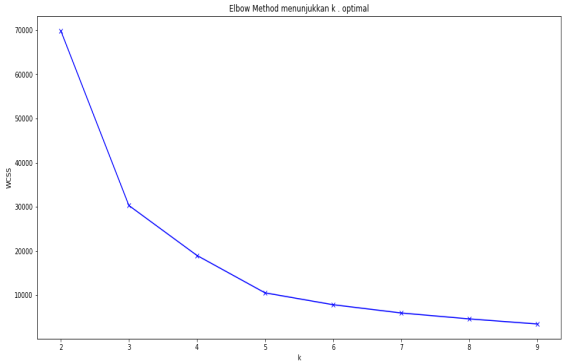

Figure 2. Elbow Graph Determination of K Values

In K3, there is a significant decrease that forms a right angle on the graph. In the Elbow method, the SSE value has decreased significantly, forming an angled shape on the graph which indicates the best number of clusters [13]. Therefore, based on the elbow method, the optimal number of clusters is 3 clusters.

b. K-Means clustering

After determining the optimal number of clusters, the next step is to group consumer data into the number of clusters obtained from the Elbow method, which is 3 clusters. The K-Means algorithm is used in Python to divide consumers into three clusters using the following command.

```
kmeanModel = KMeans(n_clusters=3, init='k-means++', max_iter=300)
```

The result of the command above is to divide each consumer into groups that can be used as material for analyzing consumer characteristics. Table 4 presents an excerpt of consumer data with K-Means and its corresponding clusters:

Table 4. Consumer and Cluster data with K-means

Institution	Educational Stage	Types of Scholarships	Category	Cluster
MAS Hikmatul amanah	S1	Dalam Negeri	SMA/MA	1
IAIN TA	S1	Luar Negeri	Perguruan tinggi	1
SMK Negeri 1 Gombong	S1	Dalam Neger	SMA/MA	3
MAS Hikmatul amanah	S1	Dalam Neger	SMA/MA	1
Universitas Indonesia	S1	Dalam Neger	Perguruan tinggi	2

The following presents a visualization of 3D K-Means data and its clusters in Figure 3 below:

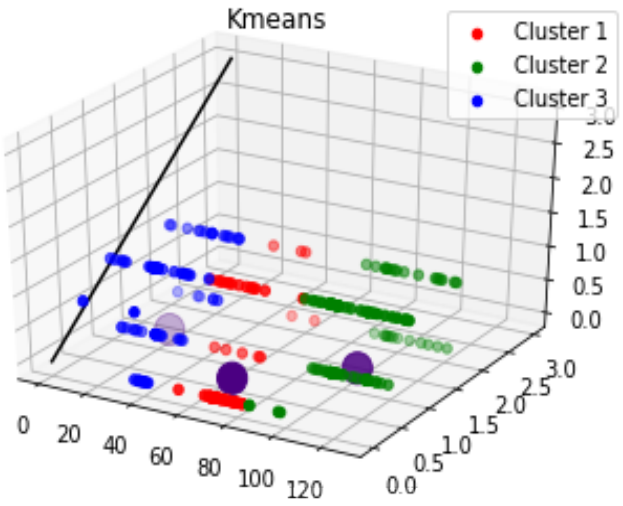

Figure 3. Visualization 3D K-means

The figure illustrates a 3D representation of Table 4, which displays three clusters: cluster 1 is represented in red, cluster 2 in green, and cluster 3 in blue. The clusters will then undergo further analysis of their characteristics. Afterward, a plot is created using the k-means method, as shown in Figure 4 below:

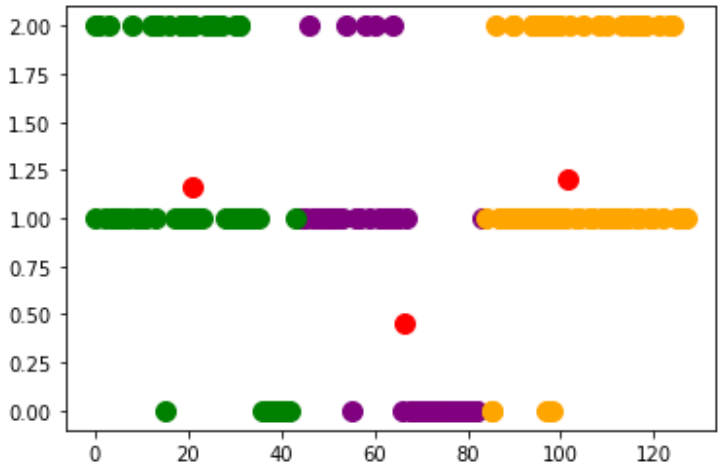

Figure 4. Visualization K-means

The plot in Figure 4 shows the result of the k-means method, where purple represents cluster 1 with 86 members, orange represents cluster 2 with 130 members, green represents cluster 3 with 64 members, and the red point is the center of the cluster. The visualization in Figure 3 is a 3-dimensional representation of the cluster distribution, as the dataset used has more than 2 attributes or variables. The purple dot in this figure represents the center of the cluster.

c. Cluster testing

The results of cluster testing using the Silhouette Coefficient are visualized in Figure 5:

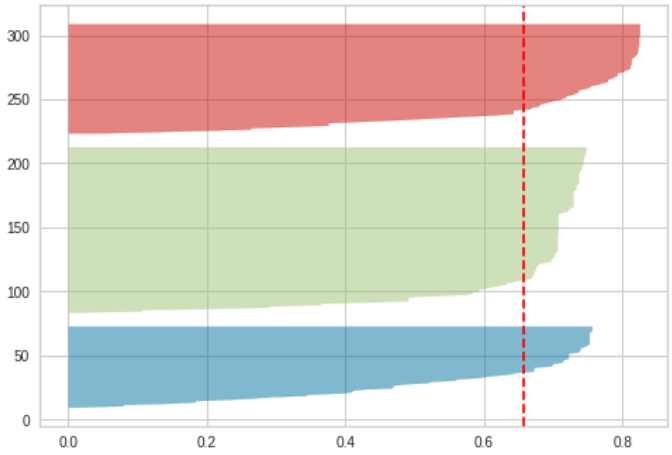

Figure 5. Visualization Silhouette Coefficient

The results of cluster testing using the Silhouette Coefficient in the K-means algorithm have produced a silhouette value of 0.657. The silhouette score in clustering ranges from -1, indicating very poor clustering, to 1, indicating very good clustering [14] This means that the quality of grouping data into one cluster using the K-means method is strong, as there is no value behind 0.

3.2. Hierarchical clustering

a. Determine the optimal number of clusters

To determine the number of clusters, the Elbow method is used, as in the previous K-means method. The sum of squared errors (SSE) for each experimental cluster is presented in Table 5 below.

Table 5. SSE Hierarchical

Cluster	SSE
K2	74645,40
K3	31703,58
K4	22479,53
K5	10796,79
K6	8646,23
K7	6448,68
K8	5497,70
K9	4787,22

After determining the number of clusters, a comparison of the differences between each cluster is performed to find the optimal cluster. The cluster that has the most significant difference is considered the optimal cluster. In Table 6, it can be seen that a significant decrease occurred at K3. Furthermore, a plot is generated, as shown in Figure 6, where the X-axis represents the number of K from experimental clusters, ranging from 2 to 9 clusters, and the Y-axis represents the SSE value obtained from each K value on the X-axis. The rate of decrease in SSE using the Hierarchical method is presented in Table 6:

Table 6. The Rate of Decline of SSE Hierarchical

Kn - Kn+1	The Rate of Decline
K2 – K3	**42941,82**
K3 - K4	9224,05
K4 – K5	11682,74
K5 – K6	2150,56
K6 – K7	2197,55
K7 – K8	950,98
K8 - K9	710,48

The following graph illustrates the optimal K (cluster) value, as shown in Figure 6:

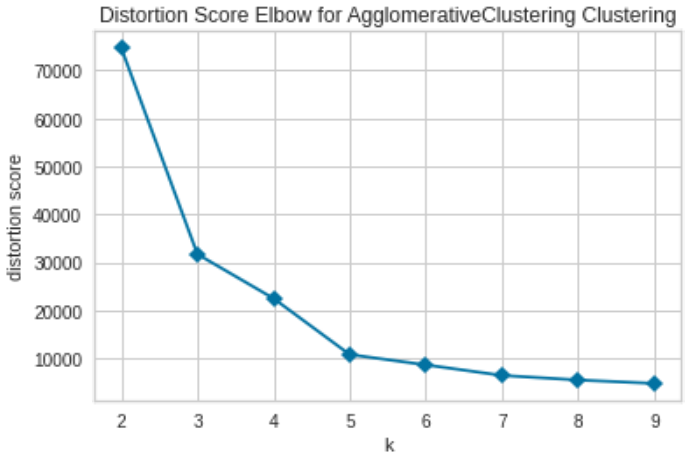

Figure 6. Graph of Optimal K Values

Based on Figure 6, it can be seen that the elbow point is formed at point 3, as there is no significant decrease anymore. Therefore, it can be concluded that the optimal number of clusters using the elbow method is 3 clusters.

b. Grouping with Average Linkage

The process of grouping consumer data using Hierarchical clustering is divided into three clusters using the Average linkage method in Python with the following command.

```
clustering = AgglomerativeClustering(n_clusters = 3, affinity = 'euclidean',
linkage = 'average')
```

The above command is used to assign each data point to a cluster and divide consumer data into groups based on the shortest distance from each cluster. It also forms a cluster dendrogram. The following is the dendrogram of the hierarchical clustering, as illustrated in Figure 7:

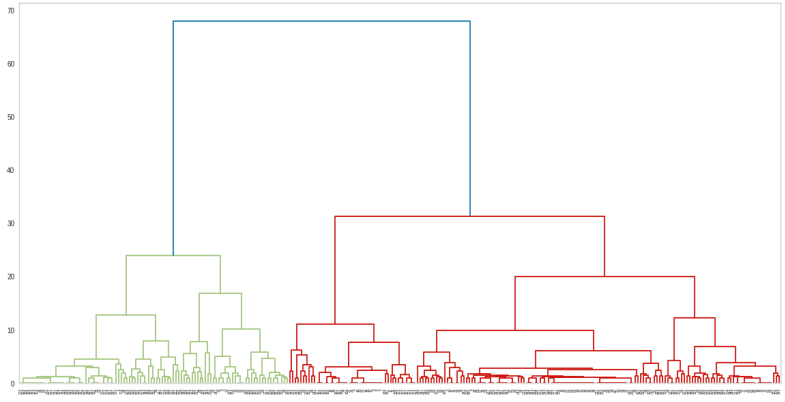

Figure 7. Dendrograms Hierarchical clustering

The following is a visualization of the plotting results from the Hierarchical clustering method, as shown in Figure 8:

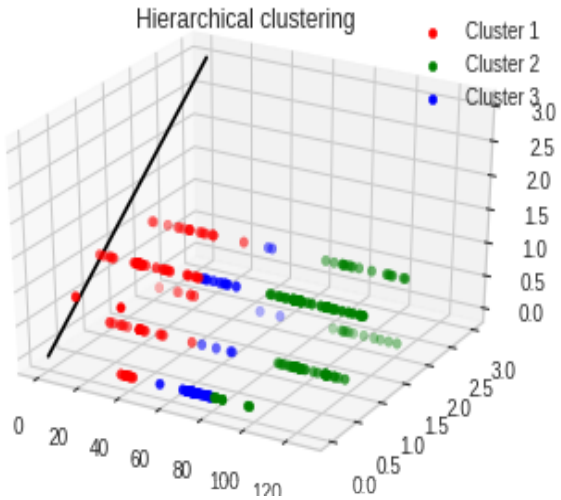

Figure 8. Visualization 3D Hierarchical

Then, a plot is generated from the Hierarchical method, as shown in Figure 9 below:

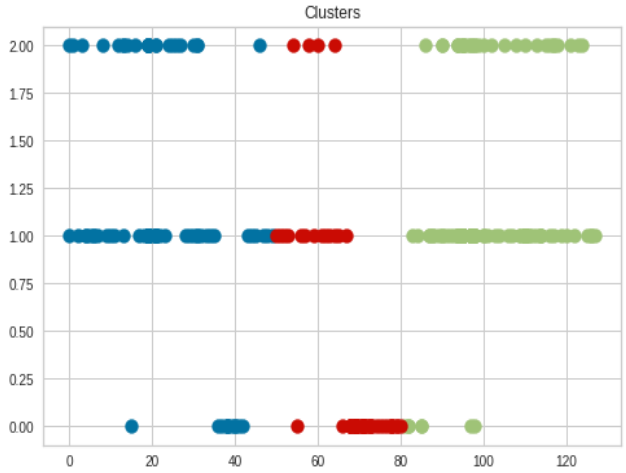

Figure 9. Visualization Hierarchical clustering

In Table 7, an excerpt of consumer data with K-Means and its corresponding clusters is presented:

Table 7. Consumer Data and its Clusters with Hierarchical Clustering

Institution	Educational Stage	Types of Scholarships	Category	Cluster
MAS Hikmatul Amanah	S1	Dalam Negeri	SMA/MA	1
IAIN TA	S1	Luar Negeri	Perguruan tinggi	1
SMK Negeri 1 Gombong	S1	Dalam Negeri	SMA/MA	3
MAS Hikmatul Amanah	S1	Dalam Negeri	SMA/MA	1
Universitas Indonesia	S1	Dalam Negeri	Perguruan tinggi	2

c. Cluster testing

The next step is to validate the hierarchical clustering method using the Silhouette Coefficient method, which is the same method used for K-means. The results of the cluster testing with the Silhouette Coefficient method on the hierarchical clustering method indicate a cluster strength level of 0.644. However, in clusters 1 and 2, there is data with a value of 0, indicating that there is incorrect data in these clusters. The results are visualized in Figure 10.

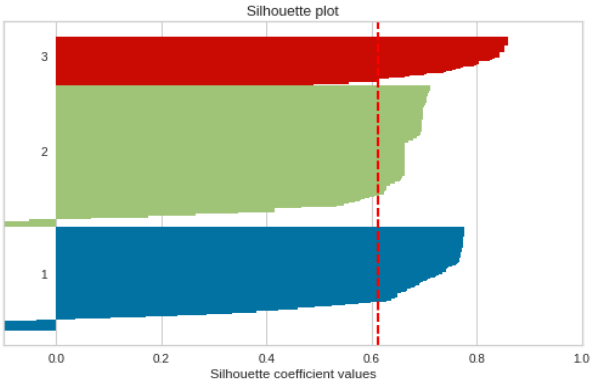

Figure 10. Silhouette Coefficient Hierarchical clustering

4. Results Analysis and Evaluation

Based on the clustering of OYI consumer data using both K-means and hierarchical clustering methods, three groups with several members who possess similar characteristics have been obtained. The clustering results are then analyzed to identify the characteristics of consumers based on institutional attributes, intended level, and types of scholarships.

4.1. K-Means

The distribution of cluster results on K-Means is described in Table 8, with the percentages shown:

Table 8. Distribution of Cluster K-means

Cluster	Total	Percentage
Cluster 1	86	31%
Cluster 2	130	46%
Cluster 3	64	23%

According to Table 8, the distribution of consumers is 23% in cluster 1, 46% in cluster 2, and 31% in cluster 3. Further details of the cluster distribution can be found in Table 9:

Table 9. Cluster Results Using The K-Means Method

Atribut		Cluster		
		1	2	3
Category	SMA / MA sederajat	42	2	10
	Perguruan tinggi	3	121	14
	Instansi Pemerintahan	2	1	29
	Swasta	12	6	9
	BUMN	1	-	-
	BUMD	4	-	-
	Individu	-	-	24
Educational Stage	S1	40	4	11
	S2	19	96	50
	S3	5	30	25
Types of Scholarships	Dalam Negeri	47	65	37
	Luar Negeri	17	65	49

Consumers in cluster 1 are primarily high school/equivalent students, with a total of 42 consumers. The most targeted level of education in this cluster is a bachelor's degree, as evidenced by the 40 consumers interested in continuing their bachelor's degree education. Consumers in cluster 2 are primarily students, as indicated by the 121 consumers coming from higher education institutions. The most targeted level of education in this cluster is a master's degree, with 96 consumers interested in pursuing this level of education. Most consumers in cluster 3 come from government agencies, with a total of 29 consumers. The majority of these consumers are interested in continuing their master's education abroad.

4.2. Hierarchical clustering

The distribution of cluster results in hierarchical clustering is presented in Table 10, with the percentages shown:

Table 10. Distribution of Data Clustering with Hierarchical Clustering Method

Cluster	Total	Percentage
Cluster 1	99	35%
Cluster 2	134	48%
Cluster 3	47	17%

According to the clustering results, 35% of consumers are in cluster 1, 48% are in cluster 2, and 17% are in cluster 3. The detailed distribution of the clusters can be viewed in Table 11:

Table 11. Cluster Results Using The Hierarchical Clustering Method

Attribute		Cluster		
		1	2	3
Category	SMA / MA sederajat	11	6	37
	Perguruan tinggi	14	121	3
	Instansi Pemerintahan	29	1	2
	Swasta	18	6	3
	BUMN	1	-	-
	BUMD	2	-	2
	Individu	24	-	-
Educational Stage	S1	12	7	36
	S2	60	97	8
	S3	27	30	3
Types of Scholarships	Dalam Negeri	41	68	40
	Luar Negeri	58	66	7

Based on the results of clustering with hierarchical clustering in Table 4.21, consumers in cluster 1 mostly come from government agencies (29 consumers) and individuals (24 consumers). Most consumers in cluster 1 (60 consumers) intend to continue their education at the master's level, and 58 consumers wish to continue their education abroad. In cluster 2, there are 121 consumers, 52 of whom come from universities. At the educational stage in cluster 2, 97 of them are pursuing postgraduate education and 68 consumers have goals of attending domestic tertiary institutions. In cluster 3, the majority of consumers come from high school/equivalent level (37 consumers). Of these consumers, 36 are interested in continuing their undergraduate education and 40 have goals of attending domestic tertiary institutions.

After conducting the analysis, the next step is to evaluate and provide recommendations for Customer Relationship Management (CRM) strategies that the marketing department of OYI can adopt to attract consumers to attend events. By comparing the results of the K-Means and Hierarchical clustering methods, it was found that the K-Means method had a higher silhouette coefficient value. Thus, the results of the K-

Means method will be used as a reference for making CRM recommendations. The comparison of the results of the K-Means and Hierarchical clustering methods can be seen in Table 12:

Table 12. Comparison of K-Means and Hierarchical Clustering Results

Index	K-Means	Hierarchical clustering
Optimum clusters	3	3
Silhouette Coefficient	0,657	0,644
Cluster 1	86	99
Cluster 2	130	134
Cluster 3	64	47

Most consumers in cluster 1 are high school/equivalent students who intend to continue their undergraduate education. In other words, the majority of members in this cluster are teenagers. A strategy that could be implemented for this group is offering OYI program promos and discounts for bringing friends, as teenagers tend to be influenced by their peers [15]. OYI can also collaborate with schools to make it easier for them to disseminate information about OYI programs and events, in order to establish long-term relationships with consumers. Additionally, OYI can focus on branding on social media platforms that target the millennial generation, by providing educational content related to the programs offered by OYI.

Cluster 2 has the highest percentage among the other clusters, and the consumers in this cluster are mostly students. To establish a relationship with this group, OYI can provide promotional information and special events through social media, and conduct surveys to gather feedback on their experiences participating in OYI activities. Strategies such as providing student-friendly pricing or discounts and offering free webinar events can also be applied to consumers in this cluster. The marketing division can also create engaging content about continuing master's education abroad, as well as information on domestic and foreign scholarships, as the most targeted level of education in this cluster is postgraduate.

improve grammar Based on Table 4.15, Cluster 3 is dominated by government agency employees who wish to develop their careers to a higher level and are more dominant in applying for education abroad. The right strategy applied to this cluster is by giving alumni testimonials, free pre-tests, and offering personal programs that can be done flexibly according to their needs. In addition, OYI can provide offers in the form of additional curriculum vitae review sessions and mentoring packages for several people who are part of a marketing strategy so as to attract more customers..

CONCLUSION

Segmentation of the characteristics of mentoring consumers is carried out using K-Means and Hierarchical Clustering. The optimal cluster obtained by the elbow method was three clusters, which were tested based on categories, intended education, and types of scholarships. Cluster 1, which is dominated by high school/vocational high school students and their equivalents, Cluster 2, which is mostly university students, and Cluster 3, which is dominated by employees of government agencies. Of the three clusters, Cluster 1 had the largest silhouette coefficient. It is hoped that by determining these resulting clusters, it can assist OYI management in determining strategies to improve Customer Relationship Management through the service targets provided. After testing, it was concluded that the more optimal results were obtained using the K-Means method with a score of 0.657, while the score of hierarchical clustering was 0.644.

REFERENCES (Min. 15 Articles)

[1] Afandi T, “Siaran Pers BONUS DEMOGRAFI 2030-2040: STRATEGI INDONESIA TERKAIT KETENAGAKERJAAN DAN PENDIDIKAN,” 2017.

[2] N. Falikhah, “BONUS DEMOGRAFI PELUANG DAN TANTANGAN BAGI INDONESIA,” 2017. [Online]. Available: https://media.neliti.com/media/publications/48298-ID-

[3] A. D. Savitri, F. Abdurrachman Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang),” 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[4] G. Febrina Wulandari, “SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM) PADA HIJAB MIULAN,” 2014, Accessed: Oct. 26, 2021. [Online]. Available: http://eprints.dinus.ac.id/5543/

[5] K. Auliasari *et al.*, “Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R,” 2019.

[6] S. N. Gama, I. Cholissodin, and M. T. Furqon, “Clustering Portal Jurnal Internasional Untuk Rekomendasi Publikasi Berdasarkan Kualitas Cluster Menggunakan Kernel K-Means,” *Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, vol. 5, no. 1, 2014, [Online]. Available: https://www.researchgate.net/publication/343876684

[7] D. Ayu, I. C. Dewi, and K. Pramita, “Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” 2019.

[8] M. Anggara, H. Sujiani, and H. Nasution, “Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness,” *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 2016.

[9] P. M. Arini and R. A. Hendrawan, “KARAKTERISTIK PELANGGAN TELEPON KABEL MENGGUNAKAN CLUSTERING SOM DAN K-MEANS UNTUK MENGURANGI KESALAHAN KLASIFIKASI PELANGGAN PERUSAHAAN TELEKOMUNIKASI (STUDI KASUS : PT. TELKOM MOJOKERTO),” 2013.

[10] A. Chusyairi and P. Ramadar Noor Saputra, “Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering,” *Telematika*, vol. 12, no. 2, pp. 139–148, Aug. 2019, doi: 10.35671/telematika.v12i2.848.

[11] A. Muhidin, “ANALISA METODE HIERARCHICAL CLUSTERING DAN K-MEAN DENGAN MODEL LRFMP PADA SEGMENTASI PELANGGAN,” *Jurnal SIGMA*, vol. 7, 2017.

[12] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, “Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah,” *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 584–594, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[13] A. Burhan, H. Kiat, Y. Azhar, and V. Rahmayanti, “Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary),” *REPOSITOR*, vol. 2, no. 7, pp. 945–952, 2020.

[14] F. Wang, H. H. Franco-Penya, J. D. Kelleher, J. Pugh, and R. Ross, “An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10358 LNAI, pp. 291–305. doi: 10.1007/978-3-319-62416-7_21.

[15] A. O. Carissa, “PENERAPAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) SEBAGAI UPAYA UNTUK MENINGKATKAN LOYALITAS PELANGGAN (Studi Kasus pada Bandung Sport Distro Malang) ANATASHA ONNA CARISSA,” Malang, 2013.