Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Identification of Public Library Visitor Profiles using K-means Algorithm based on The Cluster Validity Index

Salnan Ratih Asriningtias[1)*], Eka Ratri Noor Wulandari[2)], Myro Boyke Persijn[3)], Novita Rosyida[4)], Bayu Sutawijaya[5)]
[1,2,3,4,5)]Universitas Brawijaya
[1)]salnan@ub.ac.id, [2)]ekaratri@ub.ac.id, [3)]myrobpersijn@ub.ac.id, [4)]novitarosyida@ub.ac.id, [5)]bayu_sutawijaya@ub.ac.id

Submitted : Aug 17, 2023 | **Accepted** : Sep 16, 2023 | **Published** : Oct 1, 2023

Abstract: The existence of a public library in has a positive impact, such as increasing the literacy culture of the village community , but it still has few visitors. One of the factors that attracts visitors is the library collection. To increasing visitor interest, libraries must be able to provide book collections that can fulfill visitors' need by mapping library visitors' needs. The mapping can be done by identifying visitor profiles by grouping visitors based on the criteria of age, gender, type of visitor, and category of book library. One of the methods that can be used in the process of grouping visitors based on criteria is to use the K-Means Clustering method. Determining the number of K cluster centers at K-Means Clustering method that are not appropriate will give bad results, it is necessary to test the number of K cluster centers using the Cluster Validity index by measuring the clusters with cluster variance, within-cluster variance, and between-cluster variance. From the grouping process using K-Means Clustering with Cluster Validity index, we get 3 clusters of visitor profiles with a cluster variance value of less than 0.1. This shows that this method was able to identify the visitor profiles with high grouping accuracy values

Keywords: Between-Cluster Variance, Cluster Validity Index, Cluster Variance, K-Means Clustering, Visitor Profiles, Within-Cluster Variance

INTRODUCTION

According to statistical data from UNESCO, the reading interest of the Indonesian people is very concerning, it is only 0.001%. That means, out of 1,000 Indonesians, there is only 1 person who reads a lot (Anisa et al., 2021). This shows that there is a need to increase literacy culture in the Indonesian population to create knowledgeable population. The factor that supports the increase in literacy culture is the increasing interest of the population to read books. To increase reading interest, several programs can be carried out, one of which is building libraries in all regions as comfortable places to read, large collections of books, and offering interesting activities (Tahmidaten & Krismanto, 2020). as was done in Gampingan Village, the existence of a library in the village environment has had a very positive impact including growing interest in reading, increasing the desire to achieve higher education, the opportunity to increase income and increase awareness of the surrounding environment (Bahaudin & Wasisto, 2018).

Unfortunately, the existence of a library has not been fully utilized. The number of visitors who come is still relatively small. This is because the library collection has not met the needs of visitors. The solution that will be carried out is to increase the number of types of book collections that suit the needs of visitors. A large number of types of books requires that libraries must be able to identify what types

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

of books, can meet the needs of visitors by identifying the profile of the targeted visitors. Identification of visitor profiles is to find out what types of books are often needed by visitors or are currently trending. It can provide recommendations for procuring books, so that existing books meet visitors' need and attract visitors' interest.

To make it easier to identify visitor profiles, it is necessary to analyze the profile grouping of library visitors. The visitor profile that needs to be identified is personal factors and the history of the collection of books that have been borrowed. These personal factors include age, gender, and type of visitor. The method that can be used in the grouping process is K-means clustering (Amanda & Sitorus, 2021)(Sudrajat et al., 2022a), however, if the number of K-clusters is not correct, the results will not be optimal. To overcome this, it is necessary to test the K-Cluster using the Cluster Validity index (Widia et al., 2021). Therefore, in this study, the process of grouping visitor profiles uses the K-Menas Clustering method with the Cluster Validity Index.

LITERATURE REVIEW

Several data mining methods for clustering data are k-means (Amalina et al., 2022) and fuzzy c means (Rouza & Fimawahib, 2020). But, in process clustering k-means clustering given better results compared to fuzzy c-means (Manik et al., 2023) (Ghosh & Dubey, 2013). When using k-means clustering, the data must be in numeric format, so it is necessary to convert the data into number if the value in text format (Sudrajat et al., 2022b).

In general, there are many parameters values that will calculated in the k-means clustering process. To avoid different scale value ranges between one parameter and each other, it is necessary to normalized the data so that each parameter has the same value range scale comparison. There are many methods in the data normalization process, including zscore (Zaki et al., 2022), decimal scalling (Kusnaidi et al., 2022) and min-max (Gunadi, 2022). From several research show that min-max is able to provide the best result (Dwididanti et al., 2022) (Azzahra Nasution et al., 2019). However, in k-means clustering there are no rules for determining the number clusters, so it is necessary to carried out test analysis for different number clusters (Hou, 2019). The analysis stage to obtain the optimal number clusters can be carry out using several cluster validity index methods, including calculating distances within and between clusters (Bharill & Tiwari, 2014) and variance (Famalika & Sihombing, 2021) (Singh et al., 2017).

METHOD

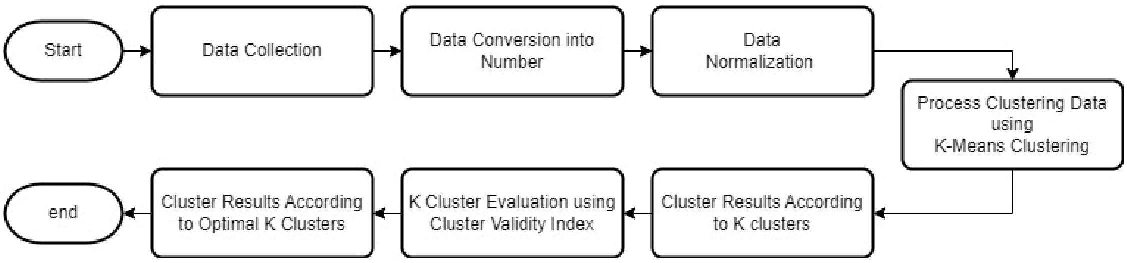

Figure 1. Flowchart of Research

Figure. 1 explains the stages of the research wich explained sequentially as follows:

Data Collection

There are two factors in the process of grouping visitor profiles, personal factors and historical factors of the types of books borrowed (Ningrum et al., 2015). Some personal factors include age, gender, and type of visitor. Visitor profile data can be seen in Table 1.

Table 1. Visitor Profile Data

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

No	Age	Gender	Visitor Type	Book Category
1	33	Female	Employee	Social
2	12	Male	Student	Science and Math
3	9	Male	Student	Science and Math
4	32	Female	Employee	Technology
5	8	Male	Student	Science and Math
6	27	Female	Employee	Other
7	33	Male	Employee	Religion
8	10	Male	Student	Science and Math
9	9	Male	Student	Other
10	8	Male	Student	Other
11	9	Male	Student	Language and dictionary
12	11	Male	Student	Other
13	9	Female	Student	Science and Math
14	28	Female	Employee	Technology
15	9	Male	Student	Other
16	10	Male	Student	Other
17	34	Female	Employee	Art and Entertainment
18	10	Male	Student	Other
19	8	Male	Student	Other
20	34	Male	Employee	Art and Entertainment
21	30	Male	Employee	Religion
22	10	Female	Student	Science and Math
23	9	Female	Student	Science and Math
24	8	Female	Student	Science and Math
25	21	Female	Student	Other
26	18	Female	Student	Language and dictionary
27	23	Female	Student	Science and Math
28	20	Female	Student	Science and Math
29	15	Male	Student	Other
30	8	Male	Student	Other
31	13	Male	Student	Science and Math
32	20	Female	Student	Science and Math
33	25	Female	Student	Science and Math
34	19	Female	Student	Science and Math
35	20	Female	Student	Other
36	22	Female	Student	Technology
37	23	Female	Student	Technology
38	20	Female	Student	Technology
39	33	Male	Employee	Social
40	16	Female	Student	Technology
41	20	Female	Student	Technology
42	17	Female	Student	Other
43	20	Female	Student	Other

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

No	Age	Gender	Visitor Type	Book Category
44	14	Female	Student	Other
45	12	Female	Student	Language and dictionary
46	21	Female	Student	Other
47	20	Female	Student	Other
48	31	Male	Employee	Technology
49	29	Male	Employee	Technology
50	34	Male	Employee	Language and dictionary
51	17	Female	Student	History and Geography
52	21	Female	Student	Other
53	17	Female	Student	Science and Math
54	23	Female	Student	Science and Math
55	34	Male	Other	Other
56	35	Female	Other	Other
57	32	Female	Other	Language and dictionary
58	24	Female	Student	Science and Math
59	24	Female	Student	Science and Math
60	22	Female	Student	Science and Math
61	22	Female	Student	Science and Math
62	17	Female	Student	Religion
63	58	Male	Other	Other
64	15	Female	Student	Literature
65	10	Female	Student	Literature
66	17	Female	Student	Science and Math
67	17	Female	Student	Art and Entertainment
68	17	Female	Student	Other
69	16	Female	Student	Social
70	21	Female	Student	Social
71	31	Female	Other	Science and Math
72	29	Female	Other	Other
73	32	Female	Other	Language and dictionary
74	28	Female	Other	History and Geography
75	9	Female	Student	Technology
76	11	Female	Student	Technology
77	10	Female	Student	Technology
78	11	Female	Student	Technology
79	30	Female	Other	Literature
80	34	Female	Employee	Other
81	35	Female	Employee	History and Geography
82	27	Female	Employee	Other
83	51	Male	Other	Other
84	34	Male	Employee	Other
85	43	Female	Other	Other
86	45	Female	Other	Other

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

No	Age	Gender	Visitor Type	Book Category
87	44	Female	Other	Philosophy and Psychology
88	52	Male	Other	Other
89	52	Male	Other	Other
90	31	Female	Other	Science and Math
91	34	Female	Other	Science and Math
92	34	Female	Other	Science and Math
93	26	Female	Other	Technology
94	28	Female	Other	Technology
95	26	Male	Other	Technology
96	20	Female	Student	Other
97	25	Female	Student	Religion
98	19	Female	Student	Other
99	34	Male	Other	Technology
100	32	Male	Other	Other
101	32	Male	Other	Other
102	33	Female	Other	Other
103	19	Female	Student	Other
104	31	Female	Other	Language and dictionary
105	34	Female	Other	Other
106	18	Female	Student	Science and Math
107	19	Female	Student	History and Geography
108	44	Male	Other	Social
109	39	Male	Other	Social
110	45	Male	Other	Religion
111	37	Male	Other	Religion
112	48	Female	Other	Religion
113	18	Female	Student	Other
114	24	Female	Other	Science and Math
115	23	Female	Other	Other
116	24	Female	Other	Language and dictionary
117	17	Female	Student	History and Geography
118	19	Female	Student	Other
119	44	Male	Other	Religion
120	43	Female	Other	Other

Data Conversion

In the clustering technique, the processed data must be in the form of numbers, so it is necessary to convert text data into numeric data. Parameter of visitor profile data can be seen in Table 2.

Table 2. Parameter of Visitor Profile

Parameter	Value	Convert Value	Total
Age	< 15	1	27
	16-25	2	43
	26-35	3	36

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Parameter	Value	Convert Value	Total
	36-50	4	10
	>50	5	4
Gender	Male	1	37
	Female	2	83
Visitor Type	Student	1	67
	Employee	2	16
	Other	3	37
Book Category	Other	1	42
	Philosophy and Psychology	2	1
	Religion	3	8
	Social	4	6
	Language and dictionary	5	3
	Science and Math	6	27
	Technology	7	17
	Art and entertainment	8	3
	Literature	9	8
	History and Geography	10	5

Data Normalization

Data normalization was carry out using min-max normalization method following the equation (1) (Nasution et al., 2019):

$$Normalized(x) = \frac{x - minValue}{maxValue - minValue} \quad (1)$$

Clustering Data

The algorithm process of clustering visitor profiles using K-Means Clustering is as follows (Ratih Asriningtias & Sonalitha, 2018):

1. Determine the number of K clusters
2. Initialization of the K cluster center value in a random way
3. Placing each data to the nearest cluster center based on distance with Euclidean following the equation (2).

   $$d(x, y) = \| x - y \|^2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

   information:
   n = data dimensions
   x_i = data in i dimensions
   y_i = data cluster center in i dimensions

4. Recalculate the cluster center with the current cluster membership by calculating the average a cluster member
5. Compare the old cluster center with the new cluster center. If the center of the cluster changes, then do from the 3rd process again until the center of the cluster does not change anymore

K Cluster Evaluation

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Testing number of K cluster centers using the Cluster Validity index by measuring the the results of clusters with variance following the equation (6) (Jusia et al., 2019), cluster variance following the equation (3), within-cluster variance (Vw) following the equation (4), and between-cluster variance (Vb) following the equation (5) (Sonalitha et al., 2020).

$$v_c^2 = \frac{1}{n_c - 1}\sum_{i=1}^{n_c}\left(d_i - \bar{d}_i\right)^2 \quad (3)$$

Information:
v_c^2 = variance at cluster c
c = number of clusters
n_c = number of data in the cluster c
d_i = data i on cluster
$\bar{d}_i$ = average of data in cluster

$$v_w = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)v_i^2 \quad (4)$$

information:
N = number of data
k = number of cluster
n_i = value of data member in cluster i

$$v_b = \frac{1}{k-1}\sum_{i=1}^{k} n_i\left(d_{ij} - \bar{d}\right)^2 \quad (5)$$

information:
k = number of cluster
d_{ij} = data j in cluster i
$\bar{d}$ = average of d_i

$$V = \frac{v_w}{v_b} \quad (6)$$

V is variance

The within-cluster variance is used to see the results of the variance in the distribution of data in a cluster. The smaller the value of the within-cluster variance, so the cluster is better. Between-cluster variance is used to see the results of the variance in the distribution of data between clusters. The greater the value of between-cluster variance, the cluster is better

RESULT

In the clustering process, the first is to convert the data into numeric data according to the convert value in Table 2, which is display in table 3

Table 3. Visitor Profile Data Conversion

No	Age	Gender	Visitor Type	Book Category
1	3	2	2	4
2	1	1	1	6
3	1	1	1	6
…	…	…	…	…
117	2	2	1	10
118	2	2	1	1
119	4	1	3	3
120	4	2	3	1

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Before the clustering process, the data needs to be normalized using equation (1). For example, for the first data on the age parameter in table 3 and according to table 2, age has a minimum value of 1 and a maximum value of 5. So the normalized value for age parameter that has value=3 is as follows:

$$Usia(1) = \frac{3-1}{5-1} = 0.5$$

Result of data normalization for Table 3 can be see in Table 4.

Table 4. Visitor Profile Normalized Data

No	Age	Gender	Visitor Type	Book Category
1	0,50	1,00	0,50	0,33
2	0,00	0,00	0,00	0,56
3	0,00	0,00	0,00	0,56
…	…	…	…	…
117	0,25	1,00	0,00	1,00
118	0,25	1,00	0,00	0,00
119	0,75	0,00	1,00	0,22
120	0,75	1,00	1,00	0,00

The next stage is to determine the number K clusters. In this research, several numbers clusters used, from 2 until 10 number clusters. For example, from data in Table 4, 3 number clusters (K=3) are determined. The data each of cluster center is taken from normalized data by random way. For example, there are 3 cluster centers that can be see in Table 5.

Table 5. Data Cluster Center at The First Iteration

No	Age	Gender	Visitor Type	Book Category
1	0,50	1,00	0,50	0,67
2	0,25	1,00	0,00	0,00
3	0,25	1,00	0,00	0,56

Placing each data to the nearest cluster center based on distance with Euclidean following the equation (2). For example, calculate the distance of the first data in Table 4 to each cluster center.

$$d(1,1) = \sqrt{(0{,}5-0{,}5)^2 + (1{,}0-1{,}0)^2 + (0{,}5-0{,}5)^2 + (0{,}33-0{,}67)^2} = 0.33$$

$$d(1,2) = \sqrt{(0{,}5-0{,}25)^2 + (1{,}0-1{,}0)^2 + (0{,}5-0{,}0)^2 + (0{,}33-0{,}0)^2} = 0.65$$

$$d(1,3) = \sqrt{(0{,}5-0{,}25)^2 + (1{,}0-1{,}0)^2 + (0{,}5-0{,}0)^2 + (0{,}33-0{,}56)^2} = 0.60$$

Selengkapnya hasil perhitungan jarak dapat dilihat pada Tabel 6.

Table 6. Distance Calculation

No	Cluster 1	Cluster 2	Cluster 3
1	0,33	0,65	0,60
2	1,23	1,17	1,03
3	1,23	1,17	1,03
…	…	…	…
117	0,65	1,00	0,44
118	0,87	0,00	0,56
119	1,23	1,52	1,54
120	0,87	1,12	1,25

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A data will become a cluster member if it has the smallest distance from the cluster center. For example, the smallest distance value at the first data was obtain in cluster1, the first data would become a member of cluster1. The entire cluster membership can be see in Table 7

Table 7. Cluster Membership

No	Cluster 1	Cluster 2	Cluster 3
1	✓		
2			✓
3			✓
…	…	…	…
117			✓
118		✓	
119	✓		
120	✓		

From Table 7, 3 clusters with different membership obtained. According to Table 8, cluster 1 has 49 members, the 1st,…, 119th, and 120th. For Cluster 2 has has 29 members, the 6th,…, 113th, and 118th. Cluster 3 has 42 members, the 2nd, 3rd,…,and 117th.

Table 8. Membership of Cluster 1

No	Age	Gender	Visitor Type	Book Category
1	0,50	1,00	0,50	0,33
…	…	…	…	…
119	0,75	0,00	1,00	0,22
120	0,75	1,00	1,00	0,00

Recalculate the cluster center by calculating the average membership of each cluster. For example calculate the cluster center from cluster 1.

$$Age_{c1} = (0{,}50 + \cdots + 0{,}75 + 0{,}75)/49 = 0.58$$
$$Gender_{c1} = (1{,}00 + \cdots + 0{,}00 + 1{,}00)/49 = 0.57$$
$$Visitor\ Type_{c1} = (0{,}50 + \cdots + 1{,}00 + 1{,}00)/49 = 0.88$$
$$Book\ Category_{c1} = (0{,}33 + \cdots + 0{,}22 + 0{,}00)/49 = 0.39$$

The cluster center results for the 2nd iteration can be see in Table 9.

Table 9. Data Cluster Center at The Second Iteration

No	Age	Gender	Visitor Type	Book Category
1	0,58	0,57	0,88	0,39
2	0,20	0,66	0,07	0,02
3	0,15	0,77	0,03	0,36

Based on table 5 and table 9, there is a change in the cluster center data, so calculations need to be carry out from calculating the distance to getting the cluster center until the data of the cluster does not change anymore.

The results of grouping visitor's profiles using the k-means clustering method for 3 clusters can be seen in Table 10.

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 10. Results of Grouping Visitor Profiles

Cluster	Total Data
Cluster 1	26
Cluster 2	33
Cluster 3	61

The characteristics of visitors obtained from the clustering results of each group or cluster can be seen in Table 11.

Table 11. Characteristics of Visitors Profile each Cluster

Cluster	Age	Gender	Visitor Type	Book Category
Cluster 1	26-35	Male	Other	Science and Math
Cluster 2	<15	Male/Female	Student	Other
Cluster 3	16-25	Female	Employee/Student	Science and Math

The k-means clustering process is carry out with several different numbers of clusters. The optimal number clusters can obtained with evaluating several different cluster numbers form 2 until 10 number cluster using the cluster validity index. As shown in Figure. 2 the cluster validity index gives the smallest value at 3 with the cluster variance value (V) is 0.000334804.

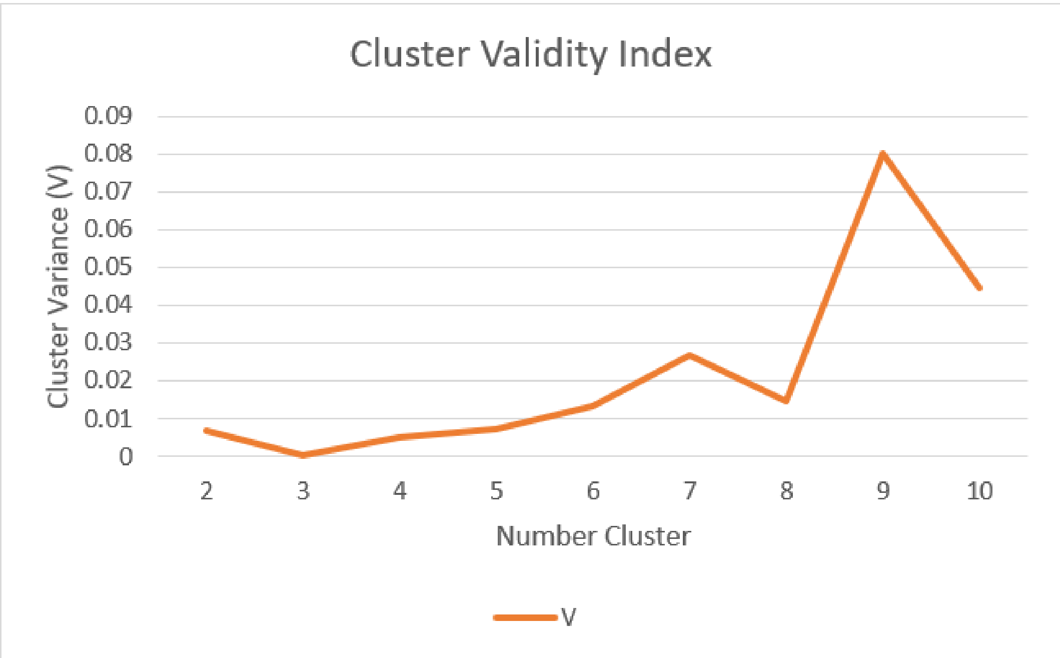

Figure 2. Cluster Validity Index of Number Cluster

DISCUSSIONS

A good cluster has a small variance (V), with small variance within a cluster (Vw) and a large variance between clusters (Vb). To prove that the method used is valid, testing is carried out for several cluster numbers which can be seen in Table 12.

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12901 p-ISSN : 2541-044X

Table 12. Cluster Number Test Results

Number of Cluster	Vw	Vb	V
2	549.122.807	813.926.805	0,674658708
3	422.207	1.261.058.308	0,000334804
4	4.634.782.609	9.483.346.661	0,488728587
5	6.938.596.491	9.806.043.504	0,707583695
6	1.106.140.351	8.415.008.517	0,131448512
7	145.840.708	5.479.814.025	0,026614171
8	9.398.230.088	6.468.399.284	1,452945261
9	3.208.108.108	4.018.124.668	0,798409301
10	1.536.607.143	3.461.124.019	0,443961885

Based on Table 5. the process of grouping visitors with 3 clusters has the smallest Vc value and the largest Vb value. This shows that the best cluster results are in the number of clusters is 3, with a cluster variance value of less than 0.1.

CONCLUSION

Based on the results of the clustering process, the profile visitor is classified into three different clusters The results of classifying visitor profiles are accurate because based on the test results, the value of a cluster variance is less than 0.1. It is necessary to improve the K-Means clustering method, especially in determining the optimal number of clusters and looking for another method to determine the initialization data of the cluster center in a random way.

REFERENCES

Amalina, T., Pramana, B. A. D., & Sari, B. N. (2022). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmiah Wahana Pendidikan*, *8*(15), 574–583. https://doi.org/10.5281/zenodo.7052276

Amanda, & Sitorus, M. V. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Konsumsi Produk Kosmetik milik PT Cedefindo. *Jurnal Ilmiah MIKA AMIK Al Muslim*, *5*(2), 63–68.

Anisa, A. R., Ipungkarti, A. A., & Saffanah, K. N. (2021). Pengaruh Kurangnya Literasi serta Kemampuan dalam Berpikir Kritis yang Masih Rendah dalam Pendidikan di Indonesia. *National Conference on Education, System, and Technology Information*, *1*(1).

Azzahra Nasution, D., Khotimah, H. H., & Chamidah, N. (2019). PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN. *Journal of Computer Engineering System and Science*, *4*(1), 2502–7131.

Bahaudin, M. S., & Wasisto, J. (2018). Peran Perpustakaan Desa Dalam Pemberdayaan Masyarakat (Studi Kualitatif Perpustakaan “Pelita” Desa Muntang). *Jurnal Ilmu Perpustakaan*, *7*(2).

Bharill, N., & Tiwari, A. (2014). Enhanced Cluster Validity Index for the Evaluation of Optimal Number of Clusters for Fuzzy C-Means Algorithm. *IEEE International Conference on Fuzzy Systems*, 1526–1533.

Dwididanti, S., Anggoro, D. A., & Sutanto, M. H. (2022). Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit. *Emitor: Jurnal Teknik Elektro*, *22*(2), 110–117. https://doi.org/10.23917/emitor.v22i2.15677

Famalika, A., & Sihombing, P. R. (2021). Implementation of K-Means and K-Medians Clustering in Several Countries Based on Global Innovation Index (GII) 2018. *Advance Sustainable Science, Engineering and Technology*, *3*(1), 0210107. https://doi.org/10.26877/asset.v3i1.8461

Ghosh, S., & Dubey, S. K. (2013). Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *IJACSA) International Journal of Advanced Computer Science and Applications*, *4*(4), 35–39. www.ijacsa.thesai.org

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12901
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Gunadi, G. (2022). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENGANALISA TRANSAKSI PENJUALAN JASA CETAK PADA UNIT PRINT ON DEMAND (POD) PERCETAKAN GRAMEDIA. *Infotech: Journal of Technology Information*, *8*(2), 117–126. https://doi.org/10.37365/jti.v8i2.148

Hou, X. (2019). A New Clustering Validity Index based on K-means Algorithm. *Journal of Physics: Conference Series*, *1187*(4), 1–7. https://doi.org/10.1088/1742-6596/1187/4/042040

Jusia, P. A., Irfan, F. M., & Kurniabudi. (2019). Clustering Data untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means. *Jurnal IKRA-ITH Informatika*, *3*(3), 75–84.

Kusnaidi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal. *Journal of Computer System and Informatics (JoSYC)*, *3*(4), 330–338. https://doi.org/10.47065/josyc.v3i4.2112

Manik, T. G., Rahayu, W. I., & Fathonah, Rd. N. S. (2023). PERBANDINGAN METODE FUZZY C-CMEANS DAN K-MEANS CLUSTERING PADA DATA PENGGUNAAN OBAT DI R.S NATIONAL HOSPITAL SURABAYA. *Jurnal Mahasiswa Teknik Informatika*, *7*(1).

Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Journal of Computer Engineering System and Science*, *4*(1), 2502–7131.

Ningrum, D. R., Saleh, C., & Permana, W. (2015). Pengaruh Kebutuhan Informasi Pemustaka Terhadap Pengembangan Koleksi Buku Tercetak pada Perpustakaan Universitas Brawijaya. *Jurnal Administrasi Publik*, *3*(5), 782–788.

Ratih Asriningtias, S., & Sonalitha, E. (2018). Aplikasi K-Means Clustering Untuk Pengelompokkan Siswa Ke Dalam Kelas Berdasarkan Nilai Akademis, Jenis Kelamin, Perilaku Dan Nama Siswa Di Sma Negeri 1 Srengat. *Seminar Nasional Sistem Informasi*, *9*(1), 1179–1187.

Rouza, E., & Fimawahib, L. (2020). Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu. *Techno.COM*, *19*(4), 481–495.

Singh, A., Vijay, V., P, R. V, & Omkar, S. (2017). Variance Based Moving K-Means Algorithm Dual-Tree Wavelet Scattering Network View project Variance Based Moving K-Means Algorithm. *IEEE International Advace Computing Conference*. https://www.researchgate.net/publication/315835294

Sonalitha, E., Zubair, A., Molyo, P. D., Asriningtias, S. R., Nurdewanto, B., Prambanan, B. R., & Mujahidin, I. (2020). Combined text mining: Fuzzy clustering for opinion mining on the traditional culture arts work. *International Journal of Advanced Computer Science and Applications*, *11*(8), 294–299. https://doi.org/10.14569/IJACSA.2020.0110838

Sudrajat, W., Cholid, I., & Petrus, J. (2022a). Penerapan Algoritma K-Means Clustering untuk Pengelompokkan UMKM Menggunakan Rapidminer. *Jurnal JUPITER*, *14*(1), 27–36.

Sudrajat, W., Cholid, I., & Petrus, J. (2022b). Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKMMenggunakan Rapidminer. *JUPITER*, *14*(1), 27–36.

Tahmidaten, L., & Krismanto, W. (2020). Permasalahan Budaya Membaca di Indonesia (Studi Pustaka Tentang Problematika & Solusinya). *Scholaria:Jurnal Pendidikan Dan Kebudayaan*, *10*(1), 22–23.

Widia, I. D. M., Rosalin, S., & Asrriningtias, S. R. (2021). Cluster Validity Index to Determine the Optimal Number Clusters of Fuzzy Clustering for Classify Customer Buying Behavior. *Journal of Development Research*, *5*(1), 7–12. https://doi.org/10.28926/jdr.v5i1.134

Zaki, A., Irwan, & Sembe, I. A. (2022). Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM). *Journal of Mathematics, Computations, and Statistics*, *5*(2), 163–176. http://www.ojs.unm.ac.id/jmathcos

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.