Analysis of the Actviness Library Members Using K-Means Clustering

Oxana Fedorova[1], Rafika Sari[1*], and Siti Setiawati[2,]

[1]*Informatics Study Program, Faculty of Computer Science, Universitas Bhayangkara Jakarta Raya, Jl. Raya Perjuangan No.81 Margamulya, Bekasi utara, 17143 Bekasi*

[*]*Email: rafika.sari@dsn.ubharajaya.ac.id*

Article Information:

Received:
2 July 2025

Received in revised form:
16 Oktober 2025

Accepted:
25 November 2025

Volume 7, Issue 2, December 2025
pp. 104 – 111

http://dx.doi.org/10.23960/jesr.v7i2.221

Abstract

*University libraries play an important role in supporting academic activities, but the trend toward digital information access has made physical services less than optimal. At Bhayangkara University Jakarta Raya, data shows a difference between the frequency of visits and book borrowing, so user segmentation is needed. This study aims to group library members based on their level of activity using the K-Means Clustering algorithm, with variables of visit frequency and borrowing. The method used is quantitative with a data mining approach, utilizing secondary data from the library system for the period May–December 2024. The analysis process includes *data preparation*, modeling using *K-Means*, and evaluation using the *Davies-Bouldin Index (DBI)*. The results show that the optimal number of clusters is three, with a DBI value of 0.628, indicating that the cluster quality is quite good. The three clusters formed are: Active Borrowers (high visits and loans), Active Non-Borrowers (high visits, low loans), and Passive Members (low visits and loans). The uniqueness of this study lies in the simultaneous combination of two user behavior variables. This segmentation is useful as a basis for developing a performance portfolio and library service strategies that are more effective and tailored to user needs.*

Keywords: *Library, data mining, K-Means Clustering, Knowledge Discovery in Database, Davies-Bouldin Index.*

I. INTRODUCTION

Libraries are one of the most important facilities in higher education institutions, serving to support learning and research activities. However, with the advancement of technology, the way students access information has shifted toward digital platforms. As a result, physical library services are no longer being utilized to their full potential. Based on data from the Library Services of the University of Bhayangkara Jakarta Raya from May to December 2024, an imbalance was found between the frequency of visits and book borrowing activities, as shown in Figure 1, indicating variations in the patterns of service utilization by users. Not all students use library facilities equally. Some students come to read, discuss, access the internet, or utilize the study room. These differences in visitation purposes indicate the importance of further analysis to understand user behavior patterns, so that the library can provide services that are more tailored to their needs.

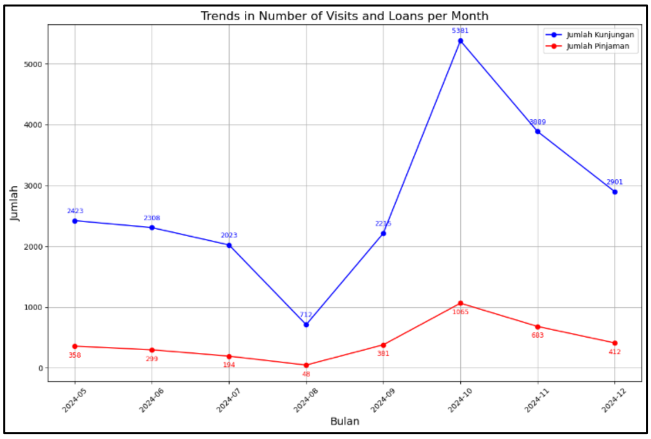

Figure 1. Book Visitation and Borrowing Patterns for May–December 2024

According to Liya Dachliyani, S.Sos., M.Pd. [1], library management includes planning, organizing, implementing, and supervising all library activities to optimize information services for users. Good management ensures that every library function runs effectively, from collection management to reader services.

One approach that can be used to understand user behavior is the clustering method, which is an unsupervised learning technique used to group data without specific labels. Data is grouped based on similar characteristics, so that similar objects are grouped into one cluster with a closer distance compared to other objects in different clusters [2].

Previous studies have shown that clustering methods, particularly K-Means, are effective in analyzing library user behavior. A study at Prima Indonesia University [3] found that data mining-based user segmentation can help evaluate user satisfaction levels and serve as a foundation for improving library service quality. Additionally, K-Means Clustering has successfully been used to group book borrowing patterns in school libraries, thereby facilitating collection management to better align with user needs [4]. Another study [5] confirms that methods like K-Means have great potential in supporting research development in the field of libraries and information science. Furthermore, a study on students [6] conducted at the Ibrahimy Library applied K-Means Clustering to analyze reading interests based on the types of books selected. The results showed that 75% of students tended to prefer fiction books, while 25% chose non-fiction books, with significant influence from external factors such as the academic environment. To enhance literacy culture, the journal recommended programs such as library visit competitions and book review contests as strategies to rebuild reading habits among students.

Based on this, this study aims to analyze the activity patterns of library members using the K-Means Clustering algorithm. This approach follows the stages of Knowledge Discovery in Database (KDD), which is the process of finding meaningful, useful, and easy-to-understand patterns from data. This study uses two main variables simultaneously, namely visit frequency and number of book loans, to produce a more comprehensive and relevant user segmentation [7].

II. Materials and Methods

A. *Stages of Knowledge Discovery in Databases (KDD)*

This study uses a quantitative method with a data mining approach through the Knowledge Discovery in Database (KDD) stages [8]. The KDD process was applied to obtain patterns of library member activity segmentation based on two main variables, namely visit frequency and number of book loans. The data used in this study is secondary data obtained from the library information system of Bhayangkara University Jakarta Raya, in the form of visit records and book loans during the period May–December 2024.

The stages in the KDD process applied in this study consist of:

1. Data Selection: The data used was selected from the library system with the criteria of the period May–December 2024. The selection of data included the number of visits and the number of member loans.
2. Data Preprocessing: This stage involves cleaning the data of empty values, duplicates, and invalid data. The data is also scaled using Min-Max Scaling to ensure that the values of each variable are within the same range, thereby improving the accuracy of the clustering results.
3. Data Mining: In this stage, the K-Means Clustering algorithm was applied to group library members into several clusters based on the variables of visit frequency and number of book loans. The computational process was performed using the Python programming language with the scikit-learn and yellowbrick libraries for visualization.
4. Evaluation: The model is evaluated to determine the optimal number of clusters using the Elbow Method, which finds the "elbow" point of the inertia value graph. Next, the quality of the clustering results is measured using the Davies-Bouldin Index (DBI), where a smaller DBI value indicates better cluster separation.
5. Knowledge Presentation: The final results of the clustering process are presented in the form of a cluster scatter plot visualization, a table of the number of members in each cluster, cluster labeling, and user segmentation interpretation. These results form the basis for compiling a library service performance portfolio.

B. *K-means algorithm*

The K-Means algorithm is one of the clustering methods used to group datasets into k clusters. Each object in the dataset is grouped based on its similarity to the cluster center (centroid) [8].

Another definition states that K-Means is a simple and effective data clustering algorithm that aims to reduce the total square error (SSE) between the data and its cluster center. This algorithm works iteratively by placing objects into the nearest cluster and updating the cluster centers until they stabilize. Because it is easy to implement and efficient, K-Means has become one of the most popular and widely used clustering methods in various applications [9].

The basic steps for the K-Means algorithm are[2]:

1. Determine the desired value of k clusters.
2. Select points or samples that will be part of the cluster at random.
3. Determine the centroid or center point of the cluster using formula (1).

$$M_k = \left(\frac{1}{n_k}\right)\sum_{i=1}^{nk} x_{ik} \qquad (1)$$

4. Calculate the square error for each cluster Ck, which is the sum of the squares of the Euclidean distances between each sample in Ck and its centroid. This error is also known as Within Cluster Variation (WCV), with the formula (2);

$$e_k^2 = \sum_{i=1}^{nk} (x_{ik} - M_k)^2 \qquad (2)$$

5. Next, the total number of errors from k-clusters is also calculated using formula (3);

$$E_k^2 = \sum_{k=1}^{k} e_k^2 \qquad (3)$$

6. Regroup all samples based on the minimum distance from each center M1, M2, …, Mk so that a new. distribution of samples according to their clusters is obtained. To obtain the new sample distribution, calculate the distance between each center point and the entire sample d(M1, x1)… d(Mk, xk). The distance between each point can be calculated using several methods, for example:

- *Euclidean Distance*

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2} \qquad (4)$$

7.Write down the results of the new cluster membership according to the results obtained in step 5.
8. Repeat step 3 several times until the total square error value decreases significantly.

C. *Davies-Bouldin Index evaluation*

The Davies-Bouldin Index (DBI), introduced by David L. Davies and Donald W. Bouldin in 1979, is a method used to evaluate cluster quality. DBI is calculated by comparing the average distance within a cluster with the distance between the nearest clusters, thereby helping to measure how well the data is grouped[10]. Once all Rij values have been calculated, the overall Davies-Bouldin Index can be calculated using formula (5):

$$DBI = \frac{1}{N}\sum_{i=1}^{N} \max_{i \neq j} R_{ij} \qquad (5)$$

III. RESULTS AND DISCUSSIONS

This study uses two main variables taken from secondary data, namely the number of visits and the number of book loans by library members during the period from May to December 2024. The data used consists of 6,426 library transaction data that have undergone cleaning and normalization processes as shown in the figure 2.

	Member ID	Frequency of Visits	Number of Loans
0	0218 12118 - rafika sari	1.098612	0.000010
1	0304027301 - dr. sugeng, sh., mh	0.000010	0.693147
2	0304027301 - sugeng, sh., mh	0.000010	0.693147
3	031503024 - adi wibowo noor fikri, s.kom, mba	0.000010	1.098612
4	0317 01064 - murti wijayanti, se., mm	1.098612	0.693147
...	...	...	...
6422	tidak diketahui - reny c. sibuea	0.000010	1.098612
6423	tidak diketahui - riyansa kanzul haqiqi, s.iip	0.000010	0.693147
6424	tidak diketahui - tutty nuryati (denda 2k)	0.000010	1.098612
6425	tidak diketahui - widya spalanzani, st., mt	0.000010	1.098612
6426	wahyu adzkia silmi - 202410325147	0.693147	0.000010

6427 rows × 3 columns

Figure 2. Modeling dataset

A. *Elbow Method*

The analysis begins with determining the optimal number of clusters using the Elbow method. The Elbow Method is applied in this study by iterating the value of K from 1 to 10 to find the optimal point, known as the "elbow point," on the inertia value graph. In this process, the KMeans object from the sklearn.cluster library was used as the clustering model, and the KElbowVisualizer from the Yellowbrick library was used to automatically visualize the inertia values without manual WSS calculations, as the inertia values were calculated directly by the KMeans function behind the scenes.

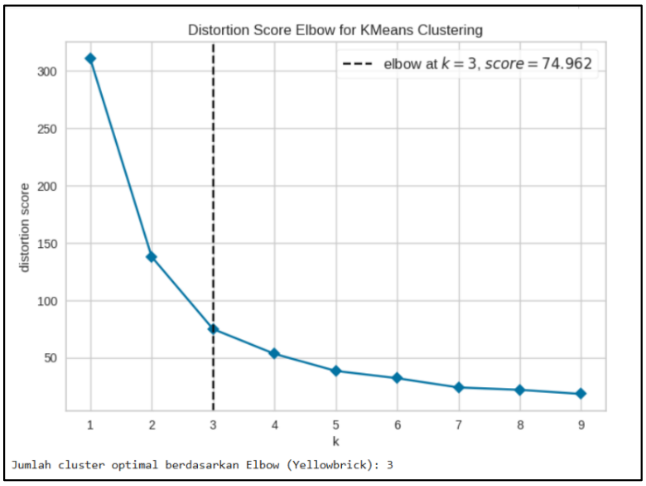

Figure 3. Elbow Method Plot Results

The resulting Elbow plot shows that the inertia value decreases significantly from K=1 to K=3, then begins to flatten out after K=3, as seen in Figure 3. The optimal point is marked by a vertical line at K=3 with a distortion score of 74.962. This point indicates the optimal number of clusters, as increasing the number of clusters beyond this value no longer results in a significant reduction in the inertia value. Therefore, the optimal number of clusters used in this study is set at 3 clusters. This result is then used for the K-Means Clustering modeling process to form 3 user segments based on visit patterns and book borrowing at the library.

B. Clustering Results

After the optimal number of clusters is obtained through the Elbow Method, the first step in the manual modeling process of the K-Means Clustering algorithm is to randomly determine the initial centroid from the previously processed dataset. In this study, three centroids were randomly selected from the dataset sample to serve as the initial center points for each cluster. Next, the Euclidean distance between each data point and the three centroids was calculated, clusters were determined based on the closest distance, and the centroid values were recalculated until the centroid values stabilized or no longer changed. The data selected as the initial centroids are shown in Table 1.

Table 1. Initial Centroid

No	Member ID	*Centroid*
1	d0236 - agus dharmanto, se, mm.	0
2	202010115069 - amryna rasyadah azahra	1
3	202210415188 - ziah febriyanti	2

The centroid used in the manual calculation process is randomly selected from the available dataset and will be used as the centroid in the first iteration. Next, the distance between each data point and the three centroids is calculated using the Euclidean Distance formula, starting from data point 1 to centroids 0, 1, and 2, up to data point n to centroids 0, 1, and 2. This calculation is performed to determine the closest distance from each data point to the existing centroids, so that each data point can be grouped into clusters based on the minimum distance.

The first data with member ID (201910415429 - Shalsa Billa Fadillah) is related to (d0236 - Agus Dharmanto, SE, MM) with centroid 0 (C_0).
The variables are described as follows:
p_1, p_2 = data values for visit frequency and data values for number of loans.
q_1, q_2 = centroid values for visit frequency and centroid values for number of loans.

$$(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

$$d(p,q) = \sqrt{(0.162 - 0.512)^2 + (0 - 0.851)^2}$$

$$d(p,q) = 0.920$$

The distance between the first data point with member ID (201910415429 - Shalsa Billa Fadillah) and (202010115069 - Amryna Rasyadah Azahra) with centroid 1 (C_1).

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

$$d(p,q) = \sqrt{(0.162 - 0.162)^2 + (0 - 0)^2}$$

$$d(p,q) = 0$$

The distance between the first data point with member ID (201910415429 - Shalsa Billa Fadillah) and (202210415188 - Ziah Febriyanti) with centroid 2.

centroid 2 (C_2).

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

$$d(p,q) = \sqrt{(0.162 - 0.323)^2 + (0 - 0)^2}$$

$$d(p,q) = 0.161$$

The results of calculating the distance of data points to each centroid using the Euclidean distance formula on the first data set, namely with member ID (201910415429 - Shalsa Billa Fadillah), produced a distance value of (C_0)= 0.920, the distance of data point 1 to (C_1)= 0, and the distance of data point 1 to (C_2)= 0.161. Next, the calculation of distance using Euclidean distance will be performed on the second data point up to the nth data point or on the dataset used.

The results of the first iteration will be used as the basis for calculating the new centroid in the second iteration, and this process will continue until the centroid values for each cluster or the membership of data in the cluster no longer change. In the second iteration, the first step is to determine the new centroid values for C_0, C_1, and C_2. These new centroid values are obtained by calculating the average of all data points that are members of each cluster based on the results of the first iteration. Next, these new centroid values will be used to calculate the Euclidean distance in the next iteration. The calculations performed are as follows:

$$C_0 = \left(\frac{x_{1,1} + x_{1,2}}{n_k}\right), \left(\frac{x_{2,1} + x_{2,2}}{n_k}\right)$$

$$C_0 = \left(\frac{0.558 + 0.512}{2}\right), \left(\frac{0.494 + 0.850}{2}\right) = \{(0.535), (0.672)\}$$

$$C_1 = \left(\frac{x_{1,1} + x_{1,2} + \ldots + x_{1,n}}{n_k}\right), \left(\frac{x_{2,1} + x_{2,2} + \ldots + x_{2,n}}{n_k}\right)$$

$$C_1 = \left(\frac{0.161 + 0.161 + \cdots + 0.161}{25}\right), \left(\frac{0 + 0 + \cdots + 0}{25}\right)$$

$$= \{(0.155), (0.027)\}$$

$$C_2 = \left(\frac{x_{1,1} + x_{1,2} + \ldots + x_{1,n}}{n_k}\right), \left(\frac{x_{2,1} + x_{2,2} + \ldots + x_{2,n}}{n_k}\right)$$

$$C_2 = \left(\frac{0.536 + 0.256 + \cdots + 0.597}{23}\right), \left(\frac{0.337 + 0 + \cdots + 0.212}{23}\right)$$

$$= \{(0.380), (0.079)\}$$

The calculation process continues until the calculation reaches convergence, or in other words, the centroid does not change with each iteration. In this study, the calculation reached convergence at the third iteration, with the results shown in Table 2.

Table 2. Results of the Last Iteration Calculation

No	Member ID	Frequency of Visits	Number of Loans	C0	C1	C2	Min	Cluster
1	201910415429 - shalsa billa fadillah	0,16155547	0	0,624	0,032	0,276	0,032	1
2	non-member - citra arindika	0,16155547	0	0,624	0,032	0,276	0,032	1
3	non-member - adzkia ramadhani ardian	0,16155547	0	0,624	0,032	0,276	0,032	1
4	20221031 5035 - gisca dwi desriyunia	0,53667565	0,33719452	0,194	0,474	0,296	0,194	0
5	202010115069 - amryna rasyadah azahra	0,16155547	0	0,624	0,032	0,276	0,032	1
	...	...	...	...	...	...	...	...
45	non-member - atsana alayya	0,16155547	0	0,624	0,032	0,276	0,032	1
46	201710415242 - raihan sidqi amrullah	0,16155547	0	0,624	0,032	0,276	0,032	1
47	d0236 - agus dharmanto, se, mm.	0,51211872	0,85098421	0,324	0,889	0,794	0,324	0
48	202110415082 - yasmin heri dharmawan	0,59782627	0,21274605	0,33	0,46	0,226	0,226	2
49	202310415225 - nur fadilah	0,16155547	0	0,624	0,032	0,276	0,032	1
50	202010325039 - crist doohan ananda mayki	0,16155547	0	0,624	0,032	0,276	0,032	1

The result of the 5th iteration centroid calculation performed using the 4th iteration Euclidean distance calculation has the same centroid result or reaches a convergent value so that the iteration can be stopped and the final value can be taken from the 4th iteration result. This can also be seen in Table 3, which compares iterations 1 to 5.

Table 3. Comparison of iterations

Cluster		C_0	C_1	C_2
Iteration 1	*Centroid p*	0.51	0.85	0.51
	Centroid q	0.16	0	0.16
Iteration 2	*Centroid p*	0.54	0.67	0.54
	Centroid q	0.16	0.03	0.16
Iteration 3	*Centroid p*	0.48	0.59	0.48
	Centroid q	0.18	0.02	0.18
Iteration 4	*Centroid p*	0.5	0.53	0.5
	Centroid q	0.18	0.03	0.18
Iteration 5	*Centroid p*	0.5	0.53	0.5
	Centroid q	0.18	0.03	0.18

C. *Computing Using Python*

After the manual calculation process is complete, the next step is to perform computations using Python on the entire available dataset. These computations are performed to process large amounts of data more efficiently and produce more accurate and objective clustering results.

The computational results show that the K-Means algorithm requires 6 iterations until the centroid position stabilizes and no longer changes. This number of iterations differs from the manual calculation results, which only require 4 iterations to achieve stability. This difference is due to the much larger amount of data used in the computational process, resulting in a more complex data distribution around the centroid and requiring more iterations until the optimal centroid position is reached.

To view the centroid values resulting from the clustering process, an additional Python script is used, as shown in Figure 3. The final centroid results from this computation also differ from the previous manual calculations, which is reasonable given the differences in data volume and distribution.

	Frequency of Visits	Number of Loans
0	0.489961	0.464384
1	0.185009	0.009963
2	0.423407	0.022223

Figure 4. The final centroid results

Overall, this computational process ensures the accuracy of the clustering results and strengthens the analysis of library user activity based on two variables: the number of visits and the number of book loans.

D. *Evaluasi Model Clustering*

The evaluation process is the final stage in research that aims to assess how well the clustering model performs. The evaluation is carried out by calculating the degree of proximity between data points, both between data points and their cluster centers (centroids), between data points within a cluster, and between clusters themselves. In this study, the Davies Bouldin Index (DBI) is used as a metric to evaluate the quality of clustering. The DBI assesses how well the clusters are separated, where the smaller or closer to zero the DBI value is, the better the quality of the model [10]. This evaluation process is carried out through a series of steps to calculate the DBI value.

The evaluation was conducted in two stages:

- Manual calculation of DBI from the results of previous clustering iterations.
    1. Calculating the Sum of Square Within Cluster (SSW) or measuring how similar or close data points in a cluster are to their centroid can also be referred to as the cohesion value. This is done using the equation (6).

    $$SSW = \frac{1}{m}\sum_{i=1}^{k} d\left(x_i\, y_j\right) \quad (6)$$

    For example, calculating SSW on the first data point relative to its cluster centroid

    $$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}$$

    $$d(p,q) = \sqrt{(0.161 - 0.177)^2 + (0 - 0.028)^2}$$

    $$d(p,q) = 0.032$$

    Next, equation 6 is calculated for all data points in the dataset. The next step is to determine the SSW value for each cluster by calculating the average distance of each cluster member.

    SSW *cluster* 0

    $$SSW_0 = \frac{0.194 + 0.157 + \cdots + 0.324}{4} = 0.187$$

    SSW *cluster 1*

    $$SSW_1 = \frac{0.032 + 0.032 + \cdots + 0.032}{32} = 0.066$$

    SSW *cluster 2*

    $$SSW_2 = \frac{0.065 + 0.081 + \cdots + 0.226}{14} = 0.134$$

    2. Calculate the Sum of Square Between Clusters (SSB) to determine how far apart the cluster centroids are, also known as the separation value, using the equation.

    $$SSB_{ij} = d(y_i, y_j) \quad (7)$$

    $$SSB_{1,0} = \sqrt{(0.496 - 0.177)^2 + (0.527 - 0.028)^2}$$

    $$SSB_{1,0} = 0.592$$

    Equation 5 was calculated for each cluster. In this study, only three clusters were used, so the calculation was performed only once, with the results shown in Table 4.

Table 4. SSB Calculation Results

SSB	C_0	C_1	C_2
C_0	0	0.592	0.471
C_1	0.592	0	0.255
C_2	0.471	0.255	0

    3. Once the SSW and SSB values have been found, the next step is to measure the ratio value to determine the comparison value of variability between values within clusters and values between clusters. The smaller the ratio value, the better. Ratio measurement can be done using the equation.

    $$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (8)$$

    $$R_{i,j} = \frac{0.187 + 0.066}{0.592} = 0.426$$

    The above calculation is performed for each cluster. In this study, three clusters were used, so the calculation was performed for each cluster and can be seen in Table 5.

Table 5. Calculation results ratio

RASIO	C_0	C_1	C_2	MAX
C_0	0	0.426	0.681	0.681
C_1	0,426	0	0,780	0.780
C_2	0.681	0.780	0	0.780

    4. The ratio value obtained is then used to calculate the DBI value using the equation 5.

    $$DBI = \frac{1}{3}(0{,}681 + 0{,}780 + 0{,}780) = \mathbf{0{,}747}$$

- DBI calculation with Python, as shown in Figure 4

```
1 # Hitung DBI score
2 dbi_score = davies_bouldin_score(scaled_data, cluster_labels)
3
4 # Tampilkan hasil
5 print(f'Davies-Bouldin Index: {dbi_score:.4f}')

Davies-Bouldin Index: 0.6282
```

Figure 5. Script to view DBI scores

The quality of the clustering results was evaluated using the Davies Bouldin Index (DBI). The computational results using Python yielded a DBI value of 0.6282, as shown in Figure 5. This value indicates that the clustering results are good, because the smaller the DBI value, the better the separation between the clusters formed.

According to reference [8], a DBI value below 0.7 is still acceptable for fluctuating public datasets, as it still indicates relatively separate clusters despite some overlap between data groups. The DBI value obtained in this study is influenced by variations in visit and borrowing data between months, as well as the presence of inactive members, which causes the distances between clusters to not be too far apart.

When compared to the results of manual evaluation using 50 sample data, a DBI value of 0.7336 was obtained. This difference in values is due to the difference in the amount of data used. The more data analyzed, the more complex the data distribution pattern relative to the centroid, so the DBI value tends to be lower and more representative.

These evaluation results also reinforce the decision to use 3 clusters in the K-Means Clustering process, as it yields the best DBI value compared to other cluster counts.

E. Discussions

This study aims to analyze the activity of members of the Bhayangkara University Jakarta Raya Library by grouping users based on their visit patterns and book borrowing habits using the K-Means Clustering algorithm. The analysis process was conducted in two stages: manual calculations on 50 data samples and computational analysis using Python on the entire dataset from May to December 2024. The clustering results were then analyzed and compared, including the number of iterations, cluster distribution, cluster labeling results, and model quality evaluation using the Davies-Bouldin Index (DBI).

In the manual calculation, the iteration process reached convergence at the fourth iteration, when the centroid values no longer changed. Conversely, the computational results using Python on the entire dataset showed six iterations until the centroid positions stabilized. This difference in the number of iterations is due to the difference in the amount of data used, where the larger the amount of data, the more complex the distribution of data relative to the centroid, so the algorithm requires more iterations to achieve stability. Additionally, the random initialization of the initial centroid in the Python computation also affects the convergence speed.

Based on the K-Means Clustering computation results, the optimal number of clusters is 3, consistent with the results of the Elbow method used previously. The distribution of members in each cluster is as follows: Cluster 0 has 738 members, Cluster 1 has 3.985 members, and Cluster 2 has 1.704 members. This distribution shows that the majority of library members belong to Cluster 1, while Cluster 0 has the fewest members. The variation in the number of members between clusters is due to differences in user activity patterns in utilizing library services, both in terms of visit frequency and book borrowing intensity.

The clustering results were then followed by labeling each cluster according to the behavioral characteristics of its members. Based on the analysis, Cluster 0 consists of members with high visit and borrowing frequencies, so it is labeled as “Active Borrowers.” Cluster 1 contains members with low visit and borrowing activities, so it is labeled as “Passive Members.” Meanwhile, Cluster 2 consists of members with high visit frequency but low book borrowing, so it is labeled as “Active Non-Borrowing Members.” This segmentation provides important information for the library to understand user behavior and serves as a basis for developing more targeted service strategies according to the characteristics of its members, as shown in Figure 6.

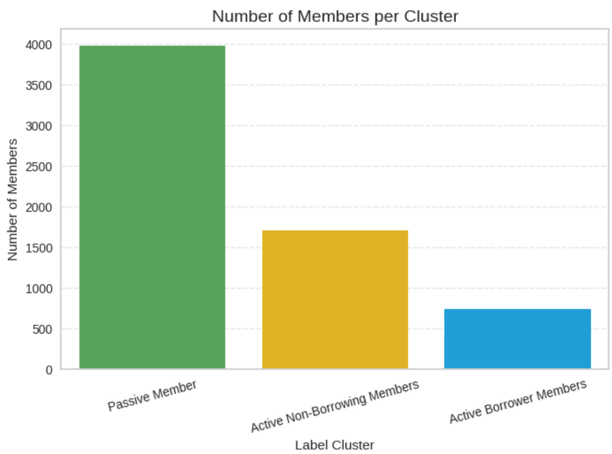

Figure 6. Histogram Clustering

IV. CONCLUSIONS

Based on the results of research conducted on the analysis of the activity of members of the Bhayangkara University Jakarta Raya library using K-Means Clustering, as well as the results of the clustering model and data labeling that has been analyzed, the following conclusions were obtained:

1. From the results of managing visit and book borrowing data over the past 8 months, a fluctuating pattern of service utilization was observed. The highest number of visits occurred in October, while August saw a significant decline. The number of book borrowings does not always follow the visit trend, as many members visit the library without borrowing books. This indicates variations in student behavior in utilizing library services;
2. The optimal number of clusters was determined using the Elbow Method, which indicated an elbow point at K=3. This result was reinforced by the

Davies-Bouldin Index (DBI) evaluation, with the lowest value of 0.6282 at K=3. Thus, library member activity data was grouped into 3 clusters;
3. The clustering analysis results show distinct characteristics in each cluster. Cluster 0 consists of members with high visit and borrowing frequencies, thus labeled as “Active Borrowers.” Cluster 1 contains members who rarely visit or borrow, labeled as “Passive Members.” Cluster 2 contains members with high visit frequency but low borrowing, labeled as “Active Non-Borrowing Members”;
4. Comparing the DBI values from manual and computational calculations shows differences in clustering quality. In manual calculations with 50 sample data, the DBI value produced is 0.747, while in computations with all data, the value is lower, namely 0.6282. This difference is due to the amount of data used, the distribution of varying values, and the sensitivity of the algorithm to the initial selection of cluster centers (centroids).

V. Acknowledgment

The author would like to thank the xxx for providing data and research facilities, as well as all those who contributed during the research process.

VI. References

[1] C. R. P. Liya Dachliyani, S.Sos, M.Pd. ; Penyunting, Drs. Ahmad Masykuri, M.M. ; Editor, *Manajemen Informasi Perpustakaan : Tuntunan Praktis Untuk Perpustakaan Umum Dan Perpustakaan Sekolah*. 2019.

[2] M. Arhami and N. Muhammad, *Data Mining - Algoritma dan Implementasi*. 2020.

[3] T. Pelayanan, P. Di, and U. Prima, “Implementasi data mining clustering dalam mengukur kepuasan terhadap pelayanan perpustakaan di universitas prima indonesia,” vol. 7, pp. 318–324, 2024, doi: 10.37600/tekinkom.v7i1.1213.

[4] D. Siburian, S. Retno Andani, I. Purnama Sari, and G. Artikel, “Implementasi Algoritma K-Means untuk Pengelompokkan Peminjaman Buku Pada Perpustakaan Sekolah Implementation of K-Means Algorithm for Clustering Books Borrowing in School Libraries,” *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i2.725.

[5] B. Lund and J. Ma, “A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering,” *Perform. Meas. Metrics*, vol. 22, no. 3, pp. 161–173, Nov. 2021, doi: 10.1108/PMM-05-2021-0026.

[6] Z. Fatah, “Analisis Pengaruh Jenis Buku Terhadap Minat Baca Mahasiswa di Perpustakaan Ibrahimy dengan Algoritma K-Means Clustering,” vol. 3, no. 1, pp. 53–65, 2025.

[7] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD Process for Extracting Useful Knowledge from Volumes of Data,” *Commun. ACM*, vol. 39, no. 11, pp. 27–34, 1996, doi: 10.1145/240455.240464.

[8] H. T. Jiawei Han, Jian Pei, *DATA MINING CONCEPTS AND TECHNIQUES*, 4th ed. Morgan Kaufmann, 2023.

[9] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. 2017.

[10] Y. Sopyan, A. D. Lesmana, and C. Juliane, “Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.