

K-Means Cluster as a Reading Interest Analysis Tool in the North Sumatra Provincial Library

Adam Amarullah Lubis^{1*}, Tengku Mohd Diansyah²

^{1,2}Universitas Harapan Medan
adamamarullah10@gmail.com^{1*}

Abstract

This study aims to apply the K-Means algorithm in analyzing students' reading interests in regional libraries in North Sumatra. As library collections and information sources increase, understanding readers' preferences has become a key element in improving the quality of library services and collections. The K-Means algorithm was chosen because of its ability to group data based on similar characteristics, thus providing deeper insight into students' reading interest patterns. The data used in this study were obtained from the visitor information database per library, which includes the number of visits, types of books frequently borrowed, duration of visits, and categories of the most popular reading materials. The data were processed through several stages, namely preprocessing to clean and align the data, clustering using the K-Means algorithm, and cluster analysis. The analysis revealed that students can be grouped into several categories based on their reading interest patterns, such as groups with high interest in scientific literature, groups that predominantly read entertainment and fiction books, and groups with low levels of reading interest. Furthermore, the results of this study also show that students' reading interests fluctuate over time. The results of this study have practical implications for library managers in developing more targeted strategies, such as literacy programs, procuring books that suit visitors' needs, and optimizing data-based services. Therefore, this study is expected to help efforts to increase the number of visits and the level of user satisfaction.

Keywords: K-Means Algorithm, Cluster, Library

1. Introduction

A library is an institution that manages and provides various sources of information, such as books, journals, magazines, and other media, which can be accessed by the public for learning, research, and recreation purposes. Libraries have an important role in supporting education and knowledge development by providing access to various physical and digital collections and services that facilitate users in utilizing the information. Libraries are one of the educational facilities to support student learning activities and play a very important role in spurring the achievement of educational goals [1]. Many libraries organize various programs and activities such as book discussions, workshops, seminars, and other cultural activities that involve the community. This makes the library a place for social interaction and shared learning [2]. Students' reading interest in libraries is often a concern because it is directly related to the quality of learning and development of science. Some of the problems faced by libraries include a decrease in the number of visitors, a lack of structured data on visitor reading interest patterns, and minimal use of technology to identify and understand students' reading interests. The decrease in the number of visitors can be caused by changes in reading habits, technological advances that offer alternative sources of information, and a lack of promotion from the library. In addition, the data held by the library is often not processed properly, so it does not provide the insight needed to understand visitor needs. The application of the K-Means algorithm can help overcome these problems. K-Means is a clustering algorithm that divides data into groups based on certain similarities. K-Means works by grouping partitions, namely dividing data into certain groups by minimizing the average distance of each data cluster [3].

The first step [4] is to collect data on library visitors, such as the number of visits, books borrowed, and borrower addresses. This data is then processed to ensure its cleanliness, followed by the application of the K-Means algorithm to group students based on their reading interest patterns. The results of this clustering are analyzed to understand the characteristics of each cluster, so that groups with high, medium, or low reading interest can be identified. Previous research shows that the product is a web application that has features that allow the system to analyze visitor and borrower data. According to the functional testing of the system, the system as a whole is successful and runs well, and the results of the user presentation show 50% Very Good, 48% Good, and 2% Less Good. The conclusion of the study is that all features can be used well on the browser. Based on this analysis, the library can formulate a more targeted strategy to increase student reading interest, such as holding promotions for certain books, book discussion events, or improving library facilities and services.

2. Theoretical Basis

2.1. Data Mining

Data mining, or data mining, is the process of extracting hidden information or patterns from large amounts of data stored in databases, data warehouses, or other storage formats. The goal is to find useful information and transform it into knowledge that can be used for various purposes, such as decision making, predictive analysis, and improving business performance. Data mining is not a new concept. Data mining is the process of exploring and analyzing large amounts of data to find meaningful and valuable patterns [5]. This concept has been around since computers were first used to store and process large amounts of data. However, the development of information technology and computerization has allowed this process to become more complex and sophisticated. Since the 1990s, data mining has evolved rapidly with the integration of statistical methods, machine learning, and artificial intelligence.

2.2. K-Means

K-Means is a non-hierarchical cluster data method that attempts to partition existing data into one or more clusters/groups. This method partitions data into clusters/groups so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into other groups. The basic concept of the K-Means algorithm begins by determining the number of clusters (k) and then randomly selecting the cluster center (centroid)[6].

2.3. Clustering

Clustering is the process of partitioning a set of data objects into subsets called clusters. Objects in a cluster have characteristics that are similar to each other and different from other clusters. In other words, Clustering is the process of grouping data into categories or groups (clusters) based on similarities. This technique is used in data analysis to find hidden patterns or relationships.

2.4. K-Means Method Calculation

Relevant data sources will be collected and prepared carefully. This data includes data on student visits to the library, the number and types of books borrowed. The use of appropriate data preprocessing techniques will be key to ensuring good data quality. The following is the data used in the study, which was taken from library data over a certain period.

Table 1: Loan Table

	Member Name	Visit	Loaning
0	Afrinaldi	1	1
1	Fathia Hanifah	1	1
2	Fika Rindani	1	1
3	Gladies Muchbita	1	1
4	Afrinaldi	1	1

1. Step1:Select Initial Centroid

We Will Choose 2 Initial Centeroids Randomly From This Data. For Example, We Choose:

- Centroid 1 (C1): Afrinaldi (1, 1)
- Centroid 2 (C2): Abdul Hamid (1, 1)

2. Step 2: Calculate Euclidean Distance

We will calculate the Euclidean distance between each data and the selected centroid. The formula for the Euclidean distance between two points (x_1, y_1) and (x_2, y_2) is:

$$Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

3. Step 3: Calculate Distance for First Data (Afrinaldi):

$$Distance C1 = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

$$Distance C2 = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

The first data is closer to C1 or C2 (since the distance is the same, we can choose either one). The second data is also the same, closer to C1 or C2.

4. Step 4: Grouping Data to Nearest Centroid

Since the distance between the data and the centroid is the same, we can randomly group the data into the centroid. For example:

- Cluster 1: Afrinaldi, Fathia Hanifah P, Fika Rindiani, Gladies Muchbita, Kaila Heri Yani
- Cluster 2: Linda Sarah Hsb, Rafli Pasya Ma'rif, Abdul Hamid, Adelia Fitri, Aliyah Hafsyahnah

5. Step 5: Recalculate the Centroid

Once the data is clustered, we recalculate new centeroids for each cluster. Since our data is identical, the new centeroids will remain the same:

- Centroid 1 (C1): (1, 1)
- Centroid 2 (C2): (1, 1)

6. Step 6: Convergence

At this stage, the centeroids do not change, so the algorithm stops here. The clustering results remain the same as those calculated in Step 4.

3. Analysis and Design

3.1. System Analysis

System analysis in this study can be defined as the process of breaking down a library information system that includes student visit and book borrowing data into its component parts. The goal is to identify and evaluate student reading interest patterns, obstacles in library services, and the needs expected by users. With this analysis, improvements to the library service system can be proposed through the application of the K-Means algorithm, which will help libraries understand student reading interests and design more effective and targeted service improvement strategies.

3.2. Problem Analysis

Problem analysis refers to the process of investigating and understanding the problem to be solved through this research. In this case, analyzing the root causes related to the decline in students' reading interest in the library and the lack of optimal utilization of visit and book borrowing data. The focus of the problem analysis is to understand the challenges and difficulties associated with grouping students' reading interest based on visit and book borrowing data. The decline in the number of library visitors can be caused by the lack of information about students' reading preferences and the lack of proper promotion. In this study, the application of the K-Means algorithm will be used to group students based on their reading interest patterns in order to identify effective solutions in increasing reading interest in the library.

3.3. System Requirements Analysis

System analysis involves various types of analysis related to the design of the system being developed as well as a thorough understanding of the needs required by an existing system. At this analysis stage, the main objective is to collect the necessary information and design a model concept that can identify student reading interest patterns in the library. The research process begins by observing the information needed in model development using the K-Means approach in data mining. The information that is the main focus includes visit data, book borrowing data. In the implementation of the student reading interest grouping system in the library, the K-Means algorithm will be used.

3.3.1. Functional Requirements

Functional requirements refer to the types of requirements that cover the operations that will be performed by the system. In addition, functional requirements also include information that must be present in the system and the results produced by the system. In the context of designing the application to be created, here are some examples of functional requirements that are needed. The system will display the flow of the K-Means model process, including collecting library visit data, borrowing books, and processing data to identify student reading interest patterns. In the results display, the system will display diagrams and plots of student groupings based on their reading interests. The system must allow the library to see groups of students with high, medium, and low reading interests.

3.3.2. Non-Functional Requirements Analysis

Non-functional requirement analysis is an analysis needed to determine the system requirement specifications. Non-functional specifications also include elements or components needed from the system being built to being implemented. In non-functional requirement analysis, hardware, software, and user analysis requirements are explained, including the following:

1. Software

The software needed to create a student reading interest grouping system using K-Means is as follows:

- a) Google Collaboratory
- b) Required Python libraries such as Pandas, NumPy, and Scikit-learn

2. Hardware

The hardware used to create a student reading interest grouping model is as follows in this case a computer or laptop Hardware specifications used:

- a) Processor: Intel(R) Core(TM) i3 CPU M350 @2.27GHz or higher
- b) Memory: 4 GB RAM or higher
- c) HDD: 500 GB or more
- d) VGA: Intel(R) HD Graphics or equivalent

3.4. Student Reading Interest Grouping System

In an effort to group students' reading interests based on data on visits and book borrowings in the library, this study designed a reading interest analysis system that uses a data mining approach with the K-Means algorithm. The development of structures and processes for grouping students' reading interests includes data analysis techniques, collecting relevant data, and creating a system structure that allows the library to produce useful information related to reading interests. The design of this system aims to see the pattern of students' reading interests so that the library can formulate a more appropriate strategy in increasing reading interests. In this section there will be stages in the model design.

3.4.1. System Analysis Stage

First, this study conducted a comprehensive system analysis to understand the various aspects involved in implementing the K-Means algorithm for clustering students' reading interests. The study identified functional requirements, which include the expected operations of

the system, as well as non-functional requirements, such as performance and scalability. This study also evaluated the data that will be used in the clustering process, including library visit data, book borrowing data, and relevant student profiles.

3.4.2. Data Mining Approach with K-means

Next, this study will design a data mining model using the K-Means algorithm as a core part of the research system. K-Means is the right choice because of its ability to group data based on certain similarities. In this context, the algorithm will be used to group students based on visiting patterns and borrowing books in the library. This study will develop a K-Means architecture that is in accordance with the characteristics of the data used, so that the grouping results can provide accurate insights into students' reading interest patterns.

3.4.3. Data Collection and Preprocessing

Relevant data sources will be collected and prepared carefully. This data includes data on student visits to the library, the number and types of books borrowed. The use of appropriate data preprocessing techniques will be key to ensuring good data quality. The following is the data used in the study, taken from library data for a certain period, for example January to December 2022, as explained in table 2.

Table 2: Number of Visits Data

o	Date	Registration Number	Bibliographic Data	Member Number	Member Name	Officer Name
1	2022-01-04 00:00:00	1C2.21441/R/2018	Salahuddin Al - Ayyubi : Riwayat hidup, legenda. dan imperium islam / John Man Cet. 3 jakarta : Pustaka Alvabet, 2018 297.98	20082700002	AFRINALDI	Maslan
2	2022-01-04 00:00:00	c3.9659/r/17	Harry Potter dan Batu Bertuah / J. K. Rowling Cet.25 Jakarta : PT. Gramedia Pustaka Utama, 2000 Fiksi Indonesia 823 Row	20011000022	FATHIA HANIFAH P	emyfitria
3	2022-01-04 00:00:00	C3.35417/R/2021	Nebula / Tere Liye JakartaGramedia Pustaka Utama : Gramedia Pustaka Utama, 2020 899.221 3	20021100022	FIKA RINDIANI	emyfitria
4	2022-01-04 00:00:00	C4.3699	Katak Keluar Tempurung : 22 Cerita yang Menghidupkan Keseharian Anda / Beny Sudarmaji Cet. 1 Jakarta : PT. Ufuk Publishing House, 2012 Fiksi Indonesia 899.221 R	17121300002	GLADIES MUCHBITA	emyfitria
5	2022-01-04 00:00:00	C3.30073/R/2019	Nagra dan aru / Ingrid Sonya Jakarta : Gramedia Pustaka Utama, 2019 Fiksi - Indonesia 899.221 3 [23]	22010400019	KAILA HERI YANI	emyfitria
6	2022-01-04 00:00:00	C5.31235/R/2019	My lady 2 / Elya Handayani Cetakan pertama Jakarta : GagasMedia, 2018 Fiksi Indonesia 899.221 3	22010400016	LINDA SARAH HSB	emyfitria
7	2022-01-04 00:00:00	C1.26421/R/2019	Pergi / Tere Liye Cetakan 8 Jakarta : Republika, 2019 Fiksi Indonesia 899.221 3 [23]	19111200018	RAFLI PASYA MAR'IF	emyfitria
8	2022-01-05 00:00:00	c3.35507/R/2021	Pengantar filsafat dari masa klasik hingga postmodernisme / Ali Maksum	22010500013	ABDUL HAMID	Maslan

o	Date	Registration Number	Bibliographic Data	Member Number	Member Name	Officer Name
Yogyakarta : Ar- Ruzz Media, 2020						
101						
9	2022-01-05 00:00:00	14.765/hd/06	Sejarah & Pemikiran pendidikan Islam Jakarta : Raja grafindo, 2004	22010500024	ADELIA FITRI	Maslan
297.73						
Love Sick / Indah Handoko						
10	2022-01-05 00:00:00	C2.35871/R/.2021	Jakarta : Elex Media Komputindo, 2020	22010500003	ALIYAH HAFSYAHNAH	emyfitria
899.221 3						

3.5. Calculation of K-Means Method

Based on Table 2, I will simplify the original data into two main attributes with For example, we will use the first 10 data as an example:

1. Number of Visits: How many times each member visited the library.
2. Number of Borrowings: How many books were borrowed by each member

Table 3: Visit Data

No.	Member Name	Visit	Loaning
1	Afrinaldi	1	1
2	Fathia Hanifah P	1	1
3	Fika Rindiani	1	1
4	Gladies Muchbita	1	1
5	Kaila Heri Yani	1	1
6	Linda Sarah Hsb	1	1
7	Rafli Pasya Ma'rif	1	1
8	Abdul Hamid	1	1
9	Adelia Fitri	1	1
10	Aliyah Hafsyahnah	1	1

1. Step1:Select Initial Centroid

We Will Choose 2 Initial Centroids Randomly From This Data. For Example, We Choose:

- a. *Centroid* 1 (C1): Afrinaldi (1, 1)
- b. *Centroid* 2 (C2): Abdul Hamid (1, 1)

2. Step 2: Calculate Euclidean Distance

We will calculate the Euclidean distance between each data and the selected centroid. The formula for the Euclidean distance between two points (x_1, y_1) and (x_2, y_2) is:

$$Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

3. Step 3: Calculate Distance for First Data (Afrinaldi):

$$Distance C1 = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

$$Distance C2 = \sqrt{(1 - 1)^2 + (1 - 1)^2} = 0$$

The first data is closer to C1 or C2 (since the distance is the same, we can choose either one). The second data is also the same, closer to C1 or C2.

4. Step 4: Grouping Data to Nearest Centroid

Since the distance between the data and the centroid is the same, we can randomly group the data into the centroid. For example:

- a. Cluster 1: Afrinaldi, Fathia Hanifah P, Fika Rindiani, Gladies Muchbita, Kaila Heri Yani
- b. Cluster 2: Linda Sarah Hsb, Rafli Pasya Ma'rif, Abdul Hamid, Adelia Fitri, Aliyah Hafsyahnah

5. Step 5: Recalculate the Centroid

Once the data is clustered, we recalculate new centroids for each cluster. Since our data is identical, the new centroids will remain the same:

- a. *Centroid* 1 (C1): (1, 1)
- b. *Centroid* 2 (C2): (1, 1)

6. Step 6: Convergence

At this stage, the centroids do not change, so the algorithm stops here. The clustering results remain the same as those calculated in Step 4.

4. Results and Discussion

4.1. Data Acquisition

This study was conducted to obtain data on student book borrowing in the North Sumatra Provincial Library, which will be used as input to analyze reading interest patterns using the K-Means algorithm. This data is expected to help libraries in increasing student engagement and reading interest. The first step taken in this study is data acquisition from library book borrowing reports, which include several important information such as borrowing date, return date, number of days late, master number, bibliographic data of borrowed books, and information related to library members. The data used in this study include variables such as: Borrowing Date, Return Date, Number of Days Late, Master Number, Bibliographic Data, Member Number, Member Name, Officer Name, and Retrieval Officer Name.

4.2. Data Preprocessing

Data preprocessing is a cleaning process to eliminate duplication and handle missing data [7]. After the book loan data is collected, the first step taken is the data preprocessing stage. Data preprocessing is carried out to prepare the book loan data from the library so that it is ready to be analyzed using the K-Means algorithm. This stage is very important because it involves the process of cleaning data, changing the appropriate format, and managing data to suit research needs. In this study, the author preprocessed the data by checking the completeness of the data, deleting invalid or empty data, and ensuring consistency in the date format and other numeric data. The processed data will then be used in the next analysis stage.

Table 4: Data Processing Result

	Registration Number	Member Number	Loan Period	Late Amount
0	1C2.21441/R/2018	20082700002	14	-2.0
1	C3.9659/R/17	20011000022	14	NaN
2	C3.354517/R/2021	20021100022	14	-56.0
3	C4.3699	17121300002	14	11.0
4	C3.30073/R/2019	22010400019	14	7.0

4.3. Data Transformation

The data transformation process includes changing the data type format to meet the research objectives in order to process the data set correctly [8]. In the process of analyzing student reading interest data, one important step is to convert categorical data such as Student ID Number or Member Number into a numeric format so that it can be processed by algorithms such as K-Means. For this purpose, we use the Label Encoding method which is part of the sklearn.preprocessing library. This method helps in converting categorical data into a numeric form that can be understood by the machine learning model.

Table 5: Encoding Result Table

	Registration Number	Member Number
0	98	1122
1	2816	989
2	2629	1035
3	3402	202
4	2470	1161

In this process, each unique value in the column that was originally a string (for example, a member code or student ID number) has been converted into a numeric form. For example, the Student ID value that was previously text such as "C3.9659/R/17" has been converted into the number 2816, and the Member ID that was originally a specific membership code is now converted into a numeric number such as 1122. Label Encoding is necessary because some machine learning algorithms, such as K-Means, can only work with numeric data. By converting this categorical data into numbers, the model can easily process and perform analysis based on these values without any problems related to the data format. The results of this encoding also show that there is no ordinal relationship between the numbers,

because the numbers only represent categories uniquely, not a specific rank or size. This transformed data is ready to be used in further processes, such as normalization and clustering.

Table 6: Result After Normalization

	Loan Period	Late Amount
0	-0.030956	0.085732
1	-0.030956	NaN
2	-0.030956	-1.335552
3	-0.030956	0.427893
4	-0.030956	0.322613

4.4. Silhouette Analysis

Silhouette Score is applied to determine the optimal number of clusters [9]. Before implementing the K-Means algorithm to group students based on book borrowing patterns, Silhouette Score analysis is performed to determine the optimal value of k in the clustering process. By using Silhouette Score, we can evaluate how well each data in the cluster fits into its cluster, as well as determine the ideal number of clusters for this dataset. The analysis starts with several different values of k (from 2 to 9), the K-Means algorithm is run for each value of k . After that, the Silhouette Score is calculated, which measures how well the data points are in their respective clusters. The Silhouette Score results are visualized in a bar graph, helping to select the k that gives the highest value (best k).

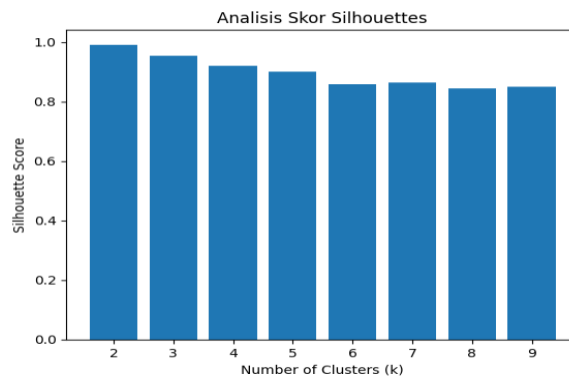


Fig. 1: Analysis Result

The bar graph shows the Silhouette Score for the number of clusters ranging from 2 to 9. Silhouette Score is a metric used to measure how well the data within a cluster is grouped, with higher scores indicating better clustering. This score ranges from -1 to 1, with values closer to 1 indicating that the data is well grouped into its respective clusters. From the graph, it appears that $k = 2$ has the highest Silhouette Score, indicating that two clusters may be the optimal choice for this dataset, as the data is more clearly separated into two distinct groups. The Silhouette Score decreases slightly for larger values of k , indicating that as the number of clusters increases, the quality of the clustering becomes less clear.

4.5. Application of K-Means

In the implementation of the K-Means algorithm to cluster students based on book borrowing patterns, the first step is to initialize the K-Means class. This class functions to build and train the K-Means model. After the model is trained, the predict method is used to predict cluster labels for new data. By using this model, the library can cluster students based on their borrowing patterns, such as borrowing duration and late returns. After clustering is complete, further analysis is carried out with dimensionality reduction using the t-SNE (t-Distributed Stochastic Neighbor Embedding) method. This dimensionality reduction aims to visualize data in two-dimensional space so that the clusters formed can be analyzed more easily. The results of the dimensionality reduction are stored in the reduced_data variable. Finally, the visualization of the clustering results is done by creating a plot. Each cluster is marked with a different color, and data points

represent students in the same cluster, based on borrowing patterns. This visualization helps distinguish groups of students who have similar book borrowing characteristics.

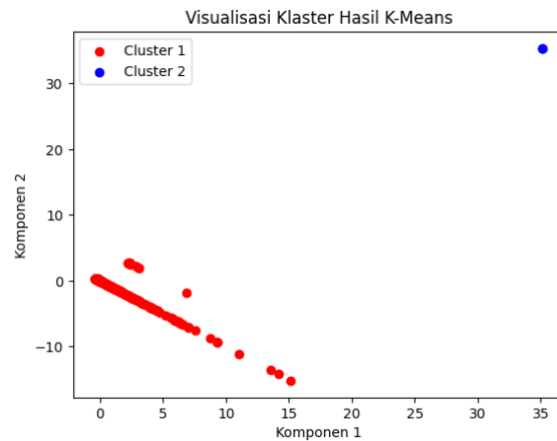


Fig. 2: Clustering Result

The visualization results of K-Means clustering after dimension reduction using PCA (Principal Component Analysis). In the graph, two clusters resulting from the clustering process are displayed in two different colors: cluster 1 (red) and cluster 2 (blue). This graph projects the data into two principal components, namely Component 1 and Component 2, which are the results of dimension reduction using PCA. The purpose of this dimension reduction is to allow us to see the distribution of clusters in two-dimensional space, making it easier to visualize the clustering patterns produced by the K-Means algorithm. In this visualization, it appears that cluster 1 (red) dominates the graph, while cluster 2 (blue) only consists of a few points. This may indicate that cluster 1 covers most of the data, while cluster 2 only contains a few data that are significantly different from the other clusters. This distribution can provide insight into the characteristics of book borrowers who are grouped into two clusters based on their behavior, such as the length of borrowing or late return of books.

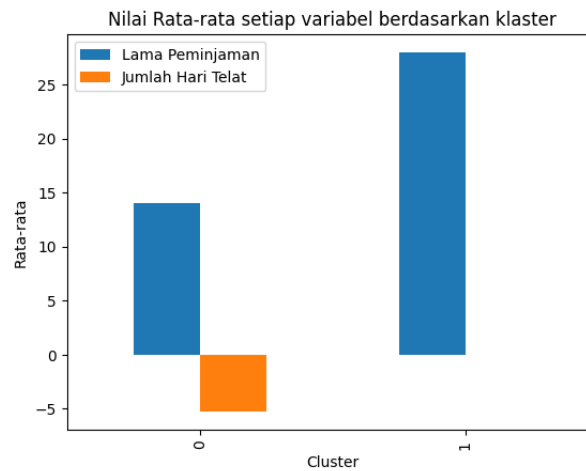


Fig. 3: Clustering Average Value

Bar graph visualization showing the average values of two variables, namely Length of Loan (blue) and Number of Days Overdue (orange), based on clusters formed by the K-Means algorithm. On the horizontal axis, there are two clusters formed, while the vertical axis shows the average of each variable. From this graph, it can be concluded that Cluster 0 has a lower average Length of Loan (around 15 days) and a negative average Days Overdue, meaning that the majority of members of this cluster return books before the deadline (not late). Cluster 1, on the other hand, has a much higher average Length of Loan (around 27 days), but no significant Days Overdue value. This could indicate that this cluster consists of borrowers who borrow books for a longer period without any recorded late returns. This graph provides important insights into the behavior of student groups in terms of borrowing and returning books. Differences in

length of loan and late returns can help libraries create more appropriate policies for each student group, such as extending the loan period for a particular cluster.

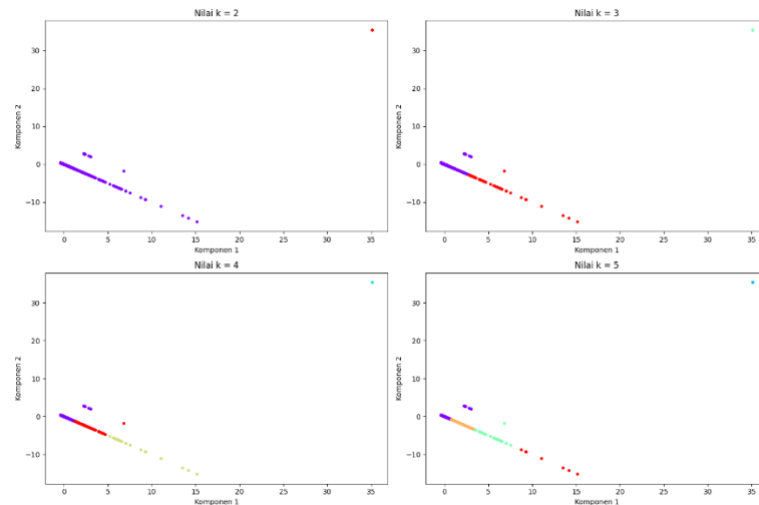


Fig. 4: Clustering Result

Visualization of K-Means clustering results with various values of k (number of clusters), namely $k=2$, $k=3$, $k=4$, and $k=5$. Each subplot shows the distribution of clusters generated by the K-Means algorithm with different numbers of clusters. This visualization was created after dimensionality reduction using PCA to display the data in two principal components (Component 1 and Component 2). Value $k=2$ The data is divided into two clusters, which is clearly seen with separate cluster distributions. One cluster has a dominant amount of data compared to the other cluster, indicating that separation into two clusters may not be enough to distinguish more specific characteristics. Value $k=3$ The addition of one more cluster shows a clearer separation between several groups of data. This cluster may provide a more detailed picture, although there is still dominance in one of the clusters.

Values of $k=4$ and $k=5$ As the number of clusters increases, the data begins to separate further into smaller groups. Some clusters may have only a few data points, indicating that with larger values of k , some clusters may not be statistically significant or are too specific. This visualization gives an idea of how the cluster structure changes as the value of k increases. By comparing the distribution of data points in each subplot, you can determine the optimal value of k , that is, the number of clusters that provides the most reasonable and meaningful separation of the data.

5. Conclusion

The results of the study indicate that libraries have an important role as providers of diverse information access to support learning and development of science. This role is becoming increasingly significant amidst the declining interest in reading among students, which is influenced by technological advances and lack of library promotion. To overcome this problem, the K-Means algorithm is used to analyze student reading interest patterns based on book borrowing and display data. A study at the North Sumatra Provincial Library by analyzing 9,134 borrowing data produced two optimal clusters based on the highest Silhouette scores. The analysis revealed that most students (75%) in the first cluster tended to return books sooner, while the rest (25%) in the second cluster borrowed books for a longer duration. The parameters used included borrowing and returning dates, borrowing deadlines, registration numbers, and bibliographic information of books to understand students' borrowing patterns and reading preferences. These findings can be the basis for designing more effective strategies in increasing reading interest and library appeal to users.

References

- [1] B. Masrufa and A. D. Ramandani, "Sinergi Kepemimpinan dan Literasi: Upaya Kepala Sekolah dalam Mengembangkan Perpustakaan Sekolah," *Irsyaduna J. Stud. Kemahasiswaan*, vol. 4, no. 1, pp. 40–55, 2024, doi: 10.54437/irsyaduna.v4i1.1550.
- [2] A. T. Permata, L. Setiawati, and L. Koerunnisa, "Analisis Penerapan Fungsi Manajemen George Robert Terry di Perpustakaan Pitomoss," *J. Librariansh. Inf. Sci.*, vol. 3, no. 2, p. 94, 2023.
- [3] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 615–620, 2024, doi: 10.36040/jati.v8i1.8360.
- [4] D. Bahtiar *et al.*, "Pengenalan Dasar Instalasi Jaringan Komputer Menggunakan Mikrotik," *Kreat. Mhs. Inform.*, vol. 2, pp. 507–518, 2021.
- [5] J. A. Noyari, A. Aprillia, R. G. Munthe, and A. Sutarnan, "Optimasi Kinerja Sistem Informasi Manajemen Kampus Menggunakan Teknik Data Mining," vol. 3, no. 1, pp. 52–63, 2024.
- [6] U. W. Latifah, S. Bahri, and M. Satriandhini, "Implementasi Algoritma K-Means Clustering untuk Strategi Promosi Kampus IBISA," *JIKO (Jurnal Inform. dan Komputer)*, vol. 8, no. 2, p. 292, 2024, doi: 10.26798/jiko.v8i2.1307.
- [7] S. Ikmi, C. Ji, P. No, and M. Kota, "ALGORITMA K-MEANS UNTUK PENINGKATAN MODEL SEGMENTASI DATA ASET TETAP PADA PT. XYZ," vol. 13, no. 1, pp. 1370–1377, 2025.
- [8] R. Nugraha, N. Suarna, I. Ali, and D. Rohman, "OPTIMASI PENGELOLAAN SAMPAH MELALUI MODEL PENGELOMPOKAN DENGAN ALGORITMA K-MEANS," vol. 13, no. 1, pp. 646–652, 2025.
- [9] F. A. Fernaldy, A. A. Arifiyanti, D. Satria, and Y. Kartika, "TRACER STUDY ALUMNI UNIVERSITAS XYZ MENGGUNAKAN ALGORITMA K-," vol. 13, no. 1, 2025.