Journal of Information Systems and Informatics
Vol. 7, No. 1, March 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i1.1037 Published By DRPM-UBD

Optimizing Library Book Circulation with K-Means Clustering: Insights from UINSU Medan

Mawar Indah Tanjung[1], Sriani[2]

[1,2]Computer Science, State Islamic University of North Sumatra, Medan, Indonesia
Email: [1]mawarindah101201@gmail.com, [2]sriani@uinsu.ac.id

Abstract

Optimal library collection management requires an understanding of book borrowing patterns to align availability with user needs. Without proper analysis, less popular books may remain in large quantities, while popular books may experience shortages. This study employs the K-Means Clustering method to group borrowed books at the Saintek UINSU Medan Library. The dataset consists of 290 loan records with attributes including book type, borrowing frequency, and the number of individuals borrowing each book. The data was converted into a numerical format and normalized using Min-Max Scaler. The Elbow Method was applied to determine the optimal number of clusters, which was found to be two. This study aims to classify books based on borrowing patterns to provide insights into library collection management. The clustering results can assist in decision-making regarding book procurement and distribution. Cluster C0 consists of popular books with high borrowing frequency and a large number of borrowers, while Cluster C1 includes books with lower borrowing rates. These findings offer a deeper understanding of borrowing trends, aiding libraries in developing acquisition strategies and organizing collections more effectively to meet user needs. These findings provide valuable insights for strategic decision-making in library collection development and maintenance, ensuring that popular books are adequately stocked while minimizing the accumulation of less-demanded titles.

Keywords: Clustering, Library, Machine Learning, K-Means Clustering

1. INTRODUCTION

Technological developments are in line with scientific advances, opening up opportunities for innovation in various fields, including Machine Learning. This technology has been applied in various aspects, such as analysis, clustering, and prediction based on predefined parameters. The Library of the Faculty of Science and Technology, State Islamic University of North Sumatra (UINSU) Medan acts as a provider of information for students regardless of gender, religion, race, age, occupation, or position. This library aims to support students in improving their knowledge and skills through access to relevant library materials. One of the main factors in achieving this goal is the availability of library materials that suit the needs of users [1].

However, in fulfilling the availability of library materials, there are still obstacles in its management. An imbalance between the number of borrowers and the stock of books available often leads to a collection vacancy if there is no addition of copies. On the other hand, the addition of book collections is not always in line with user preferences, so that some book titles that are less in demand actually experience a buildup. Therefore, research is needed to optimize library collection management with a method that can group books based on the level of user specialization. Previous studies, such as [2], have applied K-Means Clustering to group library books, but there is a lack of research focusing on real-time data analytics for dynamic clustering in academic libraries. This study aims to apply Machine Learning in grouping books based on the level of interest of users at the Scientific Library of UINSU Medan [3].

Reading interest in the library can be measured by the percentage of visitors, the number of borrowers, and the types of books that are borrowed most often. The book lending data collected every year is used as the basis for the report in determining the collection of books that need to be multiplied. Therefore, analysis of book lending transaction data is needed to obtain information related to the distribution of book borrowing intensity and the trend of reading interest of users [4].

Several previous studies have applied the K-Means Clustering method in the grouping of library books. This author conducted research at the Nurul Islam Indonesia Baru Foundation Library and found that this method can group books into three categories: (1) books that are in high demand with limited stock, (2) books that are in sufficient demand with a balanced amount of stock, and (3) books that are not in demand with abundant stock [5].

Furthermore, the research conducted by [6] obtained the results obtained between manual and RapidMiner counts from book borrower data are the same as the results that have been processed, then it was found that the number of books that were borrowed was in cluster 1 as many as 9 items, the least number of books borrowed was found in cluster 2 as many as 15 items, The number of books that are borrowed is quite large in cluster 0 as many as 12 items. The application of data mining with the k-means clustering method can be applied to book grouping so that it helps the library so that it can find out which books are often borrowed.

Based on these studies, this study aims to apply Machine Learning in grouping books based on the level of interest of users at the Scientific Library of UINSU Medan. Using the K-Means Clustering algorithm, this study will analyze book borrowing patterns based on the number of availabilities, the number of reads, and the number of borrowings. The results of this grouping are expected to provide information about the book categories that are most in demand by the academic

community of the Faculty of Science UINSU Medan and support decision-making in the procurement of library collections [7]. Thus, the *K-Means Clustering* can help libraries in minimizing variations in one cluster while maximizing variations between clusters, so that book collection management can be more optimal [8].

2. METHODS

In conducting research, the stage used is the research framework. This is done in order to prepare the stages of research. The methodology of this research is quantitative [9]. Research that tests theory by looking at the relationship between variables is called quantitative research. The purpose of this study is to implement data mining for the grouping of library loan books using K-Means Clustering (Case Study of Uinsu Medan).

Figure 1. Research Stages

The research stages, as shown in Figure 1, consist of five main steps: Planning, Data Collection, Data Analysis, Implementation, and Testing. The planning stage involves defining the research objectives and selecting the K-Means Clustering method. Data collection is carried out through literature studies, observations, interviews, and documentation studies to ensure the accuracy of the data used in the study. Data analysis includes preprocessing steps such as normalization using Min-Max Scaler to ensure that all attributes have equal weight in the clustering process

2.1. Planning

This research process begins with planning, namely determining the topic to be discussed. The topic of this research is Machine Learning for Clustering Library Loan Books Using K-Means Clustering (Case Study of the Faculty of Science Uinsu Medan).

2.2. Data Collection

Data collection in this study is carried out through four main methods, namely literature studies, observations, interviews, and documentation studies [10]. The literature study was carried out by browsing the literature related to the grouping of library books using the K-Means Clustering method to obtain a relevant theoretical foundation. Observation was carried out by directly observing the library management system at the Faculty of Science and Technology UINSU

Medan to understand the process of borrowing and the availability of books. Interviews were conducted with library management staff as the main respondents to obtain in-depth information about the book management system and the challenges faced. Meanwhile, documentation studies are used to collect technical data related to book lending by analyzing various documents and library archives. These four methods are combined to ensure the accuracy of the data used in the study.

2.3. Data Analysis

The data analysis process, at this stage, the steps carried out in detecting objects will be explained which can be seen in Figure 2.

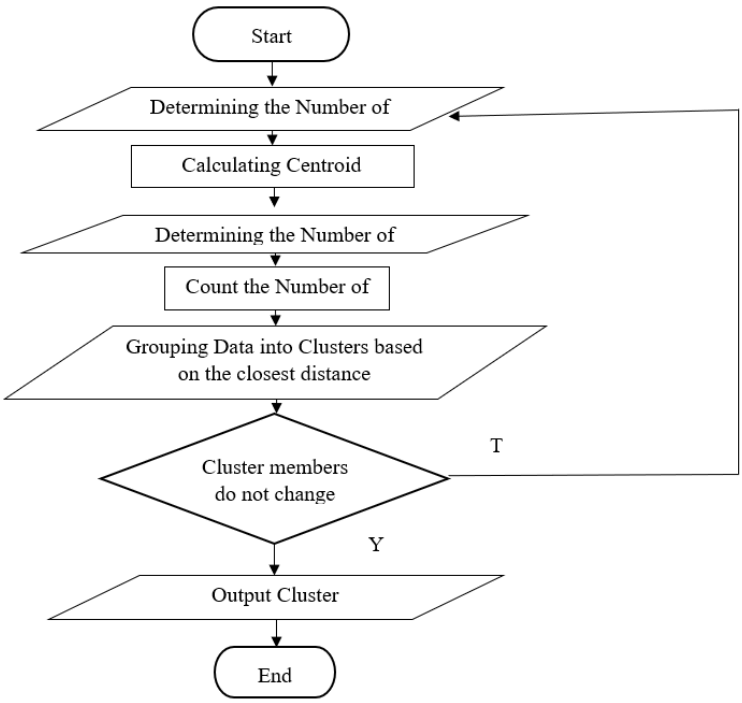

Figure 2. K-Means Method Flowchart

2.3.1. Descriptive Analysis and Data Exploration

At this stage, the focus of analysis is given on data presentation, data classification, and data compiling components. The main purpose of this stage is to obtain information and understand the content of the existing data. This stage is carried out by a literature study of several theories.

Machine learning is a branch of artificial intelligence that focuses on using algorithms and special methods to make predictions, pattern recognition, and data classification [11]. With these algorithms, complex data can be analyzed and grouped based on specific patterns, allowing libraries to identify borrowing trends and user preferences [12]. And this learning process refers to the ability of the system to perform tasks that have been learned before or to accurately generate new conclusions, based on patterns observed from previous data [13].

The K-Means algorithm is one of the partitional algorithms, because K-Means is based on determining the initial number of groups by defining their initial centroid values. The K-Means algorithm uses a process iteratively to obtain a cluster database [14]. The formula of the K-Means algorithm to calculate the distance of the data to the centeroid is used the euclidean formula, as shown in Equation 1.

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

where *d(x,y)* = Distance between the book data to the cluster central data, *xi* = Data of the ith book, and *yi* = Ith cluster central data

A university library is a library managed by a university with the aim of helping to achieve the goals of a university [15]. Higher education libraries have a role as a forum that supports teaching and learning activities, research, and service in higher education, known as the Tridharma of higher education. The university libraries in question are university libraries, institute libraries, high school libraries, polytechnics, and others. Based on these opinions, it can be concluded that a university library is a library found in a university [16]. In the research, the programming language used is python. Python is one of the programming languages that is relatively easy to learn compared to other programming languages [17].

2.3.2. Pre-Processing Data

In the data pre-processing stage, missing values were identified and imputed using the mean or median, depending on the data distribution [18]. Data transformation was performed to convert categorical data into numerical values, and normalization was applied using Min-Max Scaler to ensure that all attributes have equal weight in the cluster [19]. If the data has a normal distribution, then the missing value will be filled with the mean of the existing data [20]. However, if the data has a positively skewed or negatively skewed distribution, then the missing value will be filled with the median value of the data. By completing this data pre-processing stage, it is hoped that the data will be ready to be used for k-means clustering analysis.

2.3.3. Determining the Number of Clusters

Once the data pre-processing stage is complete, the prepared data will be used to determine the optimal number of clusters. The purpose of this step is to obtain the most appropriate k-value for cauterization analysis [21]. Determining the number of clustering can also be done automatically, this technique is carried out using the help of tools or algorithms that have been specially designed to conduct clustering analysis. One example of a tool that can be used is RapidMiner software.

2.3.4. Determining centroids and iterations with K-Means clustering analysis

After determining the optimal number of clusters, the next step is to conduct a k-means clustering analysis. In this analysis, each data will be grouped into clusters based on their distance to the center of the cluster using the Euclidean distance equation. In the k-means clustering analysis, each data will be grouped into clusters based on their distance to the nearest cluster center point. The result of this analysis is the division of data into clusters that have been formed. Each data will be included in one cluster based on its distance to the nearest cluster center point. The results of the cluster can be used for further analysis and understanding of the patterns or characteristics of the data that have been grouped.

2.4 Implementation

The application of this study began with modeling using the K-Means Clustering algorithm [22]. Based on the determination of the value of k. Sorting of calculation data. The next step is to use the K-Means Clustering category, which is the most majority.

2.5 Testing

The software that has been developed is now being tested. Testing is carried out to see if the software developed meets the demands of users by testing its functionality. A black box approach will be used to perform the testing procedure.

3. RESULTS AND DISCUSSION

The application of the K-Means method for grouping book borrowing data in libraries aims to group borrowing patterns based on certain characteristics. This research includes several main stages, starting from the collection of books borrowing data obtained from the archives of the Saintek library. This data then goes through a preprocessing process to ensure its quality, such as eliminating duplicate or irrelevant data.

The next stage is to determine the optimal number of clusters using a specific method, such as the Elbow Method or Silhouette Analysis, to ensure effective clustering results. The K-Means algorithm is then implemented to group the data into predefined clusters. After that, the system is tested to measure the accuracy and performance of the clustering results, followed by a thorough evaluation to ensure that the system can provide relevant and useful results for the library.

3.1. Data Representation

This study uses data from the archives of the Science Library of UINSU Medan which includes 290 entries, consisting of the type of book, the number of loans, and the number of borrowers. This data was analyzed using the K-Means Clustering method to group books based on borrowing patterns, thus providing insights for the management of collections and library services, as shown in Table 1. The results of the research are expected to improve operational efficiency and meet the needs of borrowers more optimally.

Table 1. Data Representation

No	Book Title	Book Type	How many times is borrowed	How Much Borrower
1	MICROORGANISM BIOLOGY & Brock	BOOK	1	1
2	(Munir Subarman) Birth History. The Development and Golden Age of Islamic Civilization. (DR. Salamah Muhammad Al-Harafi) Islamic History & Civilization Smart Book. (Dr. Muhammad Husain Mahasnah) Introduction to the Study of the History of Islamic Civilization	BOOK	1	1
3	Action Research. Operations Research in an Algorithmic Approach	BOOK	1	1

The data of this study consisted of 290 entries from the archives of the Saintek library of UINSU Medan, with three samples selected as representative to illustrate the borrowing pattern. The sample includes the type of book, the number of loans, and the number of borrowers to reflect the overall characteristics of the data. The

analysis was carried out using the K-Means Clustering method to group books based on borrowing patterns, so as to provide insights for the management and development of library services more optimally.

3.2. Drop Data

At this stage, data related to book titles will be removed from the dataset to focus the analysis on other attributes that are more relevant to the research objectives. This step is done to avoid variables that do not contribute significantly to the grouping process. It can be seen in the Table 2.

Table 2. Drop Data

It	Book Type	How many times is borrowed	How Much Borrower
1	BOOK	1	1
2	BOOK	1	1
3	BOOK	1	1

3.3. Data Transformation

Data transformation is a crucial step in analysis to standardize the format so that it can be processed numerically. In this study, the categorical data of the book type was converted into numerical values, namely books given code 1, practical work reports (KP) code 2, and thesis code 3, as shown in Table 3. This transformation simplifies data representation and improves the efficiency of the K-Means Clustering algorithm in clustering. With uniform data, the analysis becomes more optimal and accurate, ensuring that each type of book has equal weight in the clustering process.

Table 3. Categories of Book Types

Code	Book Type
1	BOOK
2	KP REPORT
3	THESIS

Once the book-type data is categorized into numerical codes, the next step is to transform the entire data to ensure format uniformity. This transformation includes all the attributes in the dataset, such as the number of borrowed and the number of individuals borrowing books, to fit the required numerical format. With structured and uniform data, the grouping process becomes more accurate, resulting in clusters that represent the actual patterns in the dataset. Additionally, these transformations also help identify relationships between attributes, providing deeper insights to support better decision-making. It can be seen in the following Table 4.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Table 4. Data Transformation

It	Book Type	How many times is borrowed	How Much Borrower
1	1	1	1
2	1	1	1
3	1	1	1

The above data is the result of a transformation, where all attributes have been converted into a uniform numerical format. The next stage is to normalize the data, which is the process of scaling the values of the attributes so that they are within a certain range, such as 0 to 1.

3.4. Data Normalization

Normalization aims to ensure that each attribute has a balanced contribution to the analysis process, so that no attribute dominates due to differences in value scales [23]. This process is particularly important in the K-Means Clustering algorithm, given that it is sensitive to scale variations between attributes. Without normalization, attributes with greater values can distort cluster formation, potentially resulting in less accurate groupings. Therefore, normalization is done so that each attribute has a uniform range of values, allowing the algorithm to work more fairly and effectively in identifying patterns in the data. In addition to improving the accuracy of grouping, normalization also plays a role in clarifying the structure of relationships between attributes and aiding in advanced analysis, such as cluster quality evaluation and data visualization. With a uniform scale, patterns of relationships between data become easier to interpret, allowing for more subtle and hidden pattern detection. One commonly used method is the Min-Max Scaler, which transforms the data values into a specific range, thus ensuring that the distribution of the data is maintained without changing its original characteristics. An example of the application of data normalization using the Min-Max Scaler can be seen in the first data in the following book lending column.

$$X' = \frac{X\text{-}Xmin}{Xmax\text{-}Xmin}$$

$$\text{D1 Book Type} = \frac{1\text{-}1}{3\text{-}1} = \frac{0}{2} = 0$$

$$\text{D1 How many times is borrowed} = \frac{1\text{-}1}{3\text{-}1} = \frac{0}{2} = 0$$

$$\text{D1 How Much Borrower} = \frac{1\text{-}1}{3\text{-}1} = \frac{0}{2} = 0$$

$$\text{D2 Book Type} = \frac{1\text{-}1}{3\text{-}1} = \frac{0}{2} = 0$$

$$\text{D2 How many times is borrowed} = \frac{1\text{-}1}{3\text{-}1} = \frac{0}{2} = 0$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

D2 How Much Borrower= $\frac{1-1}{3-1} = \frac{0}{2} = 0$

D3 Book Type= $\frac{1-1}{3-1} = \frac{0}{2} = 0$

D3 How many times is borrowed= $\frac{1-1}{3-1} = \frac{0}{2} = 0$

D3 How Much Borrower= $\frac{1-1}{3-1} = \frac{0}{2} = 0$

Table 5 shows the data on book borrowing that has gone through the normalization process. After normalization, each attribute value is adjusted within a certain range, ensuring consistency and equality of weights between attributes. This process allows for more accurate analysis and more effective grouping of data, especially when using algorithms such as K-Means.

Table 5. Data Normalization

It	Book Type	How many times is borrowed	How Much Borrower
1	0	0	0
2	0	0	0
3	0	0	0

3.5. Elbow method

The Elbow method is a technique used to determine the optimal number of clusters in a clustering analysis. This method works by calculating the inertia value or the total squared distance between the data and the cluster center for various cluster numbers. Figure 3 shows the change in inertia values as the number of clusters increases. The point at which the decrease in inertia begins to slow down or forms a sharp angle (such as an elbow) is referred to as an "elbow" and is considered the optimal number of clusters. This method helps ensure that the number of clusters selected provides a balance between model complexity and data grouping accuracy.

Based on the results of the application of the Elbow method, it was obtained that the optimal number of clusters was 2 clusters. The elbow point seen on the graph shows that the decrease in inertia values begins to slow down after the two clusters, which indicates that increasing the number of clusters further does not provide a significant improvement in the clustering results. Thus, 2 clusters were selected as the most effective number of clusters for this analysis, providing a balance between complexity and accuracy of data grouping.

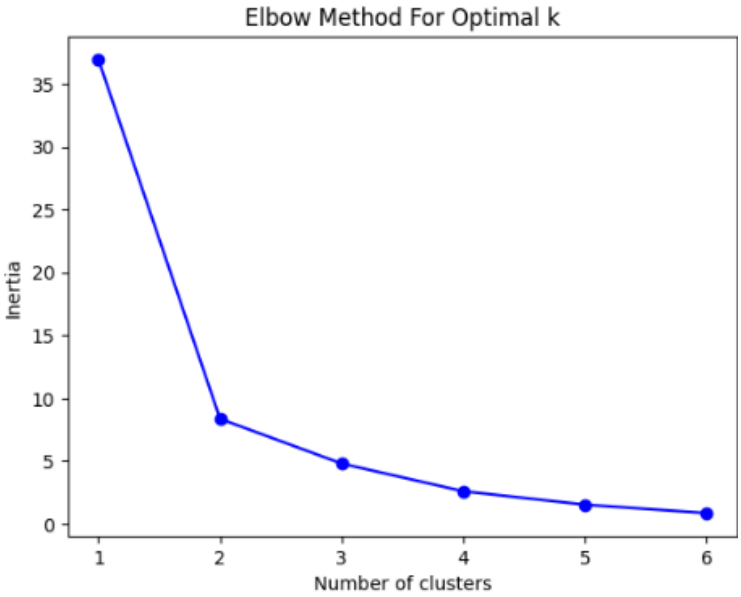

Figure 3. Elbow Method Result

3.6. Application of the K-Means Clustering method

Once the number of clusters is determined, the next step is to randomly assign the initial centroid to start the clustering process [24]. In this study, the initial centroid was selected based on the results of data normalization, where the first centroid (C0) was taken from the 19th data, and the second centroid (C1) was taken from the 7th data. Selection of a representative initial centroid is essential to ensure that the clustering process runs well and results in optimal clustering. By setting appropriate starting points, the K-Means algorithm can be more effective in finding the right cluster center and grouping data more accurately. The table below shows the initial centroid values that have been selected for each cluster. These centroid values are reflected in the following table.

Table 6. Early Centroid

Early Centroid	Book Type	How many times is borrowed	How Much Borrower
Centroid 0	1	0,28571	0,28571
Centroid 1	0	0	0

After establishing the initial centroid, the next step is to calculate the distance between each data and the centroid using the Euclidean Distance formula, as shown below. This distance calculation is important to assess the proximity of each data to a predetermined centroid, so that each data can be grouped into the most

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

appropriate cluster. By knowing the distance between the data and the centroid, the algorithm can determine which cluster has the closest center to that data, which allows the process of updating the cluster and centroid in subsequent iterations to improve the accuracy of the clustering. To calculate the distance between each data and the centroid, the following equation is used.

Calculate the distance of the nearest centroid data from 1st to 19th on centroid 0 and centroid 1 with the following attribute values:

C0 = (1; 0.28571; 0.28571)
C1 = (0; 0; 0)
The calculation of the distance between the 1st data and the nearest centroid is done with the following steps.

C0= √((1-0)^2+(0,28571-0)^2+(0,28571-0)^2)
C0= 1.078545506
C1= √((0-0)^2+(0-0)^2+(0-0)^2)
C1 = 0
…
C0= √((1-1)^2+(0,28571-0,28571)^2+(0,28571-0,28571)^2)
C0= 0
C1= √((0-1)^2+(0-0,28571)^2+(0-0,28571)^2)
C1= 1.078545506

Table 7. Iteration 1

Data	Centroid Interval 1		Cluster
	C0	**C1**	
1	1,078545506	0	1
2	1,078545506	0	1
3	1,078545506	0	1

Once the calculation using Euclidean Distance is complete, the data is grouped according to the smallest distance to the nearest centroid.

Old cluster = (0, 0

New cluster = (1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1)

Changes in data clustering indicate that the clustering process is still ongoing, so subsequent iterations are needed to update the centroid and ensure convergence.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In subsequent iterations, the new centroid is calculated based on the average of the attribute values from the data in each cluster, then the distance between each data and the centroid is updated. This process is repeated until the centroid position is stable and there is no significant change in the data grouping. Before moving on to the next iteration, the first step is to determine the new centroid by calculating the average value of each attribute from the data included in each cluster. In this case, the centroids for clusters C0 and C1 are updated based on the most recent distribution of data in their clusters. This update aims to improve the accuracy of grouping so that the clustering results generated are more representative of the patterns in the dataset.

$$C0_{\text{Book Type}} = \frac{1+1+1+\ldots+1+1}{290}$$
$$C0_{\text{Book Type}} = 0{,}993617021$$
$$C0_{\text{How many times is borrowed}} = \frac{0+0+0+\ldots+0+0}{290}$$
$$C0_{\text{How many times is borrowed}} = 0{,}030395277$$
$$C0_{\text{How Much Borrower}} = \frac{0+0+0+\ldots+0+0}{290}$$
$$C0_{\text{How Much Borrower}} = 0{,}02431634$$

C0 = (0,993617021;) 0,030395277;0,02431634

Determine the new C1 centroid based on the average value of each attribute present in cluster 1.

$$C1_{\text{Book Type}} = \frac{0+0+0+\ldots+0+0}{290}$$
$$C1_{\text{Book Type}} = 0{,}2$$
$$C1_{\text{How many times is borrowed}} = \frac{0+0+0+\ldots+0+0{,}28571}{290}$$
$$C1_{\text{How many times is borrowed}} = 0{,}025973818$$
$$C1_{\text{How Much Borrower}} = \frac{0+0+0+\ldots+0+0{,}28571}{290}$$
$$C1_{\text{How Much Borrower}} = 0{,}023376545$$

C1 = (0,2;)0,025973818;0,023376545

The new centroids acquired can be seen in Table 8, which shows the average value of each attribute for each cluster after the update is made. This table depicts the updated centroid positions, which will later be used for later iterations in the data grouping process.

Table 8. New Centroid iteration 1

Centroid	Book Type	How many times is borrowed	How Much Borrower
C0	0,993617021	0,030395277	0,02431634
C1	0,2	0,025973818	0,023376545

This calculation continues until the 3rd iteration stage because there are no further changes in the data grouping, the iteration process is considered complete. At this stage, the K-Means algorithm has reached convergence, which means that the centroid position and data grouping have stabilized. This final result indicates that the clusters formed are optimal, with data grouped based on the characteristics that have been identified during the analysis process. With the achievement of convergence, the K-Means algorithm successfully groups data according to existing patterns, providing a clearer and more structured understanding of data distribution. Where the results of the cluster can be concluded as shown in Table 9.

Table 9. Cluster

Number of Clusters	
C0	C1
232	58

The results of the clustering analysis show that Cluster C0 consists of 232 data points, representing popular books with high borrowing frequency and a large number of borrowers. Cluster C1 consists of 58 data points, representing books with lower borrowing rates. These findings provide valuable insights for library collection management, ensuring that popular books are adequately stocked while minimizing the accumulation of less-demanded title Here is the Python code to get the final centroid after the K-Means algorithm is executed. Centroid is the central point of each cluster calculated based on the average of the data positions in that cluster.

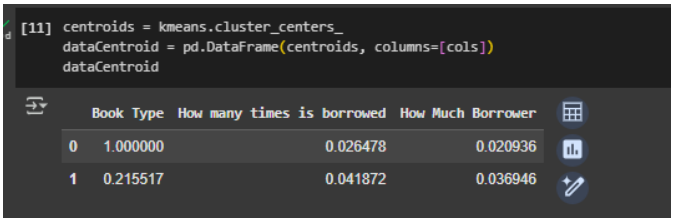

Figure 4. Late Centroid

Figure 4 shows the results of the centroid calculation in the K-Means algorithm after the clustering process is carried out. The table displays the centroid coordinates for the two clusters formed based on the attributes of book type, borrowing frequency, and number of borrowers. Cluster 0 has a centroid value of

1,000 on the book type attribute, indicating the dominance of one book category in this cluster. In contrast, Cluster 1 has a centroid value of 0.215, which indicates a more diverse variety of book types. In addition, the values on the attributes of borrowing frequency and number of borrowers indicate that Cluster 1 has a higher borrowing rate and number of borrowers than Cluster 0. The results of this grouping provide insight into the pattern of book borrowing in the library and can be the basis for optimizing the management of library collections and services.

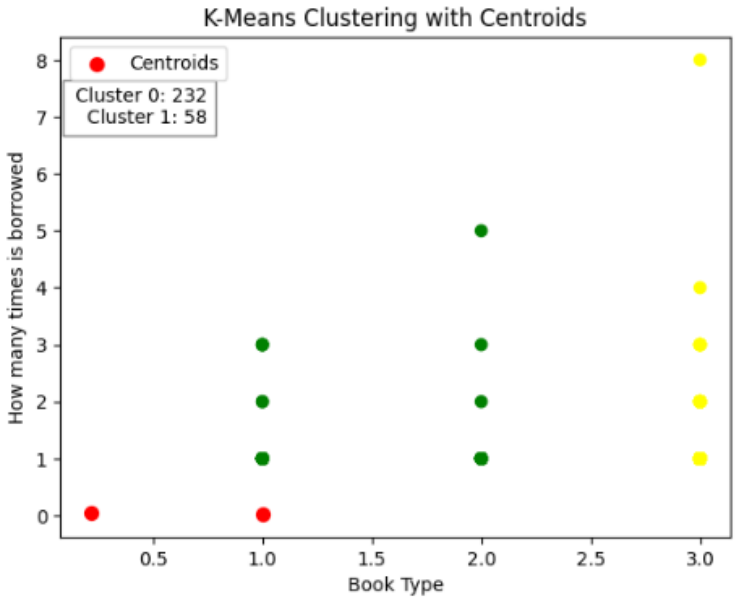

Figure 5. Scatter Plot Results

Figure 5 shows each data point that has been grouped by the K-Means algorithm, with a different color for each cluster. In addition, the centroid of each cluster is also displayed as a red dot on the plot to show the average position of the data in that cluster.

```
[18] from sklearn.metrics import mean_squared_error, davies_bouldin_score
     dbi = davies_bouldin_score(X, kmeans.labels_)
     print("Davies-Bouldin Index Value:", dbi)

     Davies-Bouldin Index Value: 0.42826312793996585
```

Figure 6. DBI Value

DBI provides an indication of how well clusters have been formed by considering the proximity of data in a single cluster and the separation between clusters. The lower the DBI value, the better the quality of the clustering produced. Compared to hierarchical clustering, K-Means offers faster computation and scalability for

large datasets, making it more suitable for library book clustering. However, hierarchical clustering may provide better interpretability for smaller datasets, as noted in [25]. The application of K-Means Clustering in this study has successfully identified borrowing patterns, which can be used to optimize library collection management.

3.7. Discussion

The application of K-Means Clustering in analyzing the book borrowing patterns at the Science Library of UINSU Medan has provided valuable insights into how the library resources are utilized. The dataset, which originally consisted of 290 entries, was methodically cleaned, transformed, and normalized before applying clustering techniques. This rigorous data preprocessing ensured that the clustering process yielded accurate and meaningful patterns.

Initially, categorical data such as book types were transformed into numerical representations to facilitate mathematical processing. With three primary types—books, practical work reports (KP), and theses—coded as 1, 2, and 3 respectively, the data was further standardized through normalization. This step was critical in neutralizing the scale of different attributes, thereby preventing any single feature from dominating the clustering outcome. The normalization using the Min-Max Scaler brought all variables into a comparable range, typically between 0 and 1, ensuring each attribute contributed equally to the distance calculations in the clustering process.

The Elbow Method was employed to determine the optimal number of clusters, and the analysis revealed that two clusters were ideal for representing the underlying borrowing patterns in the dataset. The sharp change in inertia at two clusters indicated the point where increasing the number of clusters would not significantly enhance the model's accuracy, making it a balanced choice between simplicity and explanatory power.

Once the number of clusters was established, the K-Means algorithm was executed starting with two randomly selected centroids. The algorithm iteratively recalculated distances between each data point and the centroids using the Euclidean Distance formula, updating cluster assignments based on proximity. This iterative refinement continued until convergence was achieved in the third iteration, signifying that the clusters had stabilized and no further changes in data groupings were occurring.

The final result revealed two distinct clusters: Cluster C0 with 232 entries and Cluster C1 with 58 entries. Cluster C0 predominantly represented popular book types with higher borrowing frequencies and greater numbers of borrowers. This

suggests that the majority of library users gravitate towards a specific category of books, possibly general academic references or frequently assigned texts. Meanwhile, Cluster C1, which encompassed books with lower borrowing activity, likely includes specialized, less in-demand titles.

The differentiation between clusters offers strategic value for library management. By identifying which books fall into the high-demand cluster, the library can prioritize budget allocation, increase stock for popular titles, and ensure they are readily available to users. Conversely, books in the low-demand cluster can be evaluated for relevance, potential weeding, or better promotional strategies to increase awareness among students and faculty.

A further breakdown of the final centroid values supports this interpretation. Cluster 0 had a centroid book type value close to 1.000, indicating a high concentration of a single book category—most likely traditional textbooks. Its low values for borrowing frequency and borrower count suggest that while the book type is consistent, individual titles may not be borrowed repeatedly, indicating a wide but shallow use pattern. On the other hand, Cluster 1, with a centroid book type value around 0.215, displayed more variety in book types and relatively higher values for both borrowing frequency and borrower count. This could imply a niche group of highly used resources, perhaps core materials for certain disciplines or programs.

Visual representations like the scatter plot (Figure 5) and the Davies-Bouldin Index (DBI) value further validate the effectiveness of clustering. The scatter plot distinctly shows how data points cluster around the centroids, with minimal overlap, which is indicative of good cluster separation. Meanwhile, the low DBI value signifies compact and well-separated clusters, affirming the robustness of the K-Means approach in this scenario.

Moreover, comparing the K-Means algorithm with hierarchical clustering, as briefly noted, highlights its advantages in handling large datasets due to computational efficiency and scalability. While hierarchical clustering might offer more intuitive grouping for smaller data samples, K-Means proves more practical for the extensive records managed in a university library context.

The study demonstrates that K-Means Clustering is a powerful tool for understanding borrowing patterns in academic libraries. By converting raw transactional data into actionable insights, library administrators can optimize resource allocation, improve user satisfaction, and streamline collection management. These results not only enhance operational efficiency but also support strategic decision-making for future acquisitions and library service development.

4. CONCLUSION

This study concludes that book grouping in the Scientific Library of UINSU Medan can be analyzed effectively using the K-Means Clustering method. Of the total 290 data collected, the main attributes used included the type of book, the frequency of borrowing, and the number of individuals who borrowed. To ensure that the data can be processed numerically, the book type is converted to a numerical format (Book=1, KP Report=2, Thesis=3) and normalized using the Min-Max Scaler to equalize the scale between attributes. The determination of the optimal number of clusters was carried out by the Elbow method, which resulted in two main clusters, providing segmentation in accordance with the characteristics of the analyzed data.

The grouping results show that the main factors affecting book borrowing include the type of book, the frequency of borrowing, and the number of individuals who borrow. Cluster C0 represents a more popular group of books, with a high frequency of borrowing and the number of borrowers, while Cluster C1 shows a category of books that are less in demand with a lower borrowing rate. These findings provide insights that can be used in the management of library collections, especially in decision-making related to the provision and distribution of books, so that library services can be more optimal in meeting the needs of borrowers. Future studies could explore the integration of real-time data analytics for dynamic clustering or the application of these findings in other library contexts, such as public or corporate libraries.

REFERENCES

[1] A. Febriyanto, S. Achmadi, and A. P. Sasmito, "Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan Itn Malang," *J. Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 61–70, 2021.

[2] 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.

[3] A. Anggi Riyanto, D. Daryanto, and G. Abdurrahman, "Text Mining Untuk Clustering Buku Di Perpustakaan Menggunakan Metode K-Means," *Natl. Multidiscip. Sci.*, vol. 1, no. 6, pp. 835–845, 2022, doi: 10.32528/nms.v1i6.239.

[4] I. A. Nur Afifah and H. Nurdiyanto, "Data Mining Clustering Dalam Pengelompokan Buku Perpustakaan Mengunakan Algoritma K-Means," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 3, pp. 802–814, 2023, doi: 10.29100/jipi.v8i3.3891.

[5] H. Ani, D. Nofriansyah, and I. Mariami, "Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering," *J. CyberTech*, vol. 1, no. 1, pp. 1–12, 2021, doi: 10.53513/jct.v1i1.315

[6] Yulia and M. Silalahi, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *Indones. J. Comput. Sci.*, vol. 10, no. 1, 2021, doi: 10.33022/ijcs.v10i1.3008.

[7] I. Padiku and A. Lahinta, "Penerapan Clustering K-Means Untuk Mendukung Pengelolaan Koleksi Pada Perpustakaan Fakultas Teknik Universitas Negeri Gorontalo," *J. Tek.*, vol. 20, no. 1, pp. 54–62, 2022, doi: 10.37031/jt.v20i1.206.

[8] N. N. Hasanah and A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 300–311, 2022, doi: 10.47233/jteksis.v4i2.499.

[9] R. D. Syahputra *et al.*, "Metode Penelitian Kuantitatif, Kualitatif, dan R&D," *Pers. Ed.*, vol. 4, no. 1, p. 601, 2019.

[10] M. Waruwu, "Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan," *J. Ilm. Profesi Pendidik.*, vol. 9, no. 2, pp. 1220–1230, 2024, doi: 10.29303/jipp.v9i2.2141.

[11] N. R. Muntiari and K. H. Hanif, "Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning," *J. Ilmu Komput. dan Teknol.*, vol. 3, no. 1, pp. 1–6, 2022, doi: 10.35960/ikomti.v3i1.766.

[12] M. Ula, A. Zulfikri, A. F. Ulva, and R. A. Rizal, "Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 336–348, 2023, doi: 10.33022/ijcs.v12i1.3162.

[13] A. Fathurohman, "Machine Learning Untuk Pendidikan: Mengapa Dan Bagaimana," *J. Inform. dan Teknol. Komput.*, vol. 1, no. 3, pp. 57–62, 2021, doi: 10.55606/jitek.v1i3.306

[14] S. Gustin, W. Ramdhan, and W. M. Kifti, "Teknik Data Mining Menggunakan Metode K-Means Untuk Mengcluster Dan Pencarian Buku Di Perpustakaan Daerah Kabupaten Asahan," *JUTSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, no. 3, pp. 195–204, 2022, doi: 10.33330/jutsi.v2i3.1901.

[15] I. W. Antasari and A. Noegroho, "Pengembangan Koleksi Berbasis Akreditasi Prodi Di Perpustakaan Iain Purwokerto," *Publ. Libr. Inf. Sci.*, vol. 3, no. 1, p. 12, 2019, doi: 10.24269/pls.v3i1.1666.

[16] M. D. N. H. Muhammad Dhio Nathama Harahap, M. A. Muhamad Alda, A. M. Abdul Malid, Y. F. H. Yardina Fauziah Harahap, F. Faradilla, and F. R. Fitri Rahmayani, "Efektivitas Pembuatan Sistem Informasi Manajemen Perpustakaan SMPN 1 Pantai Labu Berbasis Digital oleh KKN 110 UINSU," *Reslaj Relig. Educ. Soc. Laa Roiba J.*, vol. 6, no. 4, pp. 1555–1562, 2024, doi: 10.47467/reslaj.v6i4.765.

[17] Adawiyah Ritonga and Yahfizham Yahfizham, "Studi Literatur Perbandingan Bahasa Pemrograman C++ dan Bahasa Pemrograman Python pada Algoritma Pemrograman," *J. Tek. Inform. dan Teknol. Inf.*, vol. 3, no. 3, pp. 56–63, 2023, doi: 10.55606/jutiti.v3i3.2863.

[18] D. A. A. Prakash, "Pre-processing techniques for preparing clean and high-quality data for diabetes prediction," *Int. J. Res. Publ. Rev.*, vol. 5, no. 2, pp. 458–465, 2024, doi: 10.55248/gengpi.5.0224.0412.

[19] M. Berg, "Predictive Maintenance of Power Grids: Clustering Analysis with KMeans and Hierarchical Methods," 2024.

[20] G. Chao, Y. Jiang, and D. Chu, "Incomplete Contrastive Multi-View Clustering with High-Confidence Guiding," *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 10, pp. 11221–11229, 2024, doi: 10.1609/aaai.v38i10.29000.

[21] P. Hou, L. Zhou, and Y. Yang, "Density clustering method based on k-nearest neighbor propagation," *J. Phys. Conf. Ser.*, vol. 2858, no. 1, 2024, doi: 10.1088/1742-6596/2858/1/012041.

[22] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Ann. Data Sci.*, vol. 11, no. 5, pp. 1525–1544, 2022, doi: 10.1007/s40745-022-00428-2.

[23] M. Al Amin and D. Juniati, "Math Unesa," *J. Ilm. Mat.*, vol. 9, no. 2, pp. 437–446, 2021, doi: 10.59934/jaiea.v4i1.696

[24] A. I. Ramadhan, P. D. Atika, and K. F. Ramdhania, "Analisis Clustering K-Means untuk Pemetaan Tingkat Pengangguran Terbuka di Provinsi-Provinsi Indonesia Tahun 2013-2023," vol. 5, no. 2, pp. 109–122, 2025.

[25] K. Sugahara and K. Okamoto, "Hierarchical matrix factorization for interpretable collaborative filtering," *Pattern Recognit. Lett.*, vol. 180, pp. 99–106, 2024, doi: 10.1016/j.patrec.2024.03.003.