

Application Of The K-Means Clustering Algorithm For Data Collection And Grouping Of Reading Monitoring In The Literacy Program In One Of The Public Elementary Schools In Bandung

Chicha Wiarsa¹, Lilis Emalia², Egi Badar Sambani³, Adi Raharjo⁴
STMIK Mardira Indonesia, Bandung ^{1,2,3,4}

Email: chichawiarsa99@gmail.com¹, lilis@stmik-mi.ac.id², egi.badars@stmik-mi.ac.id³,
egi.badars@stmik-mi.ac.id⁴

Abstract

This study employs the K-Means Clustering algorithm within a data-gathering and monitoring framework for student literacy initiatives at a public primary school in Bandung. The research used a descriptive quantitative methodology, employing data from student reading activities, including the quantity of books read and reading comprehension scores. The K-Means algorithm analyzes this data to categorize children into three reading levels: high (Grade A), moderate (Grade B), and low (Grade C). The computation technique employs two variables (x and y) denoting the number of books read and the corresponding comprehension scores. The study determined the final centroids for each cluster as follows: C1 (12.5, 88.75) for high ability, C2 (7.33, 75.0) for moderate ability, and C3 (3.0, 55.0) for poor ability. After two cycles, the clustering outcomes stabilized, indicating no further member reassignments among groups. Of the 10 students assessed, the algorithm categorized four into the high cluster, three into the intermediate cluster, and three into the low cluster. The findings indicate that the K-Means algorithm effectively classifies student literacy data in an objective, quantifiable manner. The execution of this algorithm helps educators track literacy progress, categorize abilities, and develop more targeted instructional strategies.

Keywords : *K-Means Clustering, Reading Literacy, Student Clustering, Information Systems, Data Analysis*

INTRODUCTION

Primary education is essential in developing pupil literacy, which underpins academic and cognitive growth. In Indonesia, literacy initiatives are a central focus for improving reading proficiency; however, their effectiveness depends on accurate data collection and oversight.

Field evidence indicates that many schools continue to rely on manual record-keeping, resulting in imprecise data that is difficult to assess. A public elementary school in Bandung is one of the institutions confronting these difficulties. The school presently documents student reading activities using traditional methods, resulting in insufficient data on reading frequency, the number of books read, and the

categories of reading materials. The lack of student classification according to reading habits or literacy outcomes impedes the mentoring process, rendering it ineffective and poorly focused.

Information technology provides a pertinent approach to tackle these challenges. An excellent strategy is data mining with the K-Means Clustering algorithm—an unsupervised learning method adept at categorizing data based on analogous attributes. Utilizing this algorithm on student literacy data enables the school to discern reading trends, delineate student competence levels, and acquire more objective information for mentoring initiatives.

Rosário et al. (2025) define health literacy as the capacity to obtain, comprehend, evaluate, and

utilize health information for informed decision-making (PLOS ONE). In the realm of military education, Annas & Wahab (2023) and Umargono et al. (2020) underscore that media literacy involves the capacity to identify information risks, comprehend media functions, and critically assess digital content (Scientific Bulletin). Consequently, numerous contemporary studies conclude that literacy is a multifaceted skill, vital for preparing individuals to confront increasingly complex information and technological challenges.

K-Means is an unsupervised learning technique that categorizes data into clusters based on the similarity of attributes among data points. This technique is widely used by researchers across diverse domains owing to its simplicity and efficiency in managing large datasets. A fundamental attribute of K-Means is its iterative process that minimizes the distance between data points and their centroids, continually adjusting them until stability is attained. (Ahmed et al., 2020) assert that "traditional clustering algorithms such as K-means continue to be pertinent despite the advent of graph-based and deep clustering methodologies, particularly for large-scale and interpretable categorization tasks" (Abbas et al., 2020).

Laravel is a PHP framework built on the MVC (Model-View-Controller) architecture, designed to streamline web application development with its expressive, beautiful syntax. Laravel theoretically integrates object-oriented programming principles, dependency management, and contemporary design patterns to establish a framework conducive to structured, maintainable, and large-scale software

development. (Chong, 2021) assert that "Laravel is one of the most efficient object-oriented frameworks for PHP web application development, providing logical organization through the MVC model and enabling maintainable, scalable, and reusable code architecture" (Sinaga & Yang, 2020).

React JS is a JavaScript library created by Facebook for constructing dynamic and responsive user interfaces (UI) in component-oriented web applications. React JS enables developers to build easily adaptable UIs via the Virtual DOM, thereby improving performance and user experience (Ikotun et al., 2021).

MySQL is an open-source relational database management system (RDBMS) widely used by developers to store, manage, and retrieve data in online applications (Ashabi et al., 2020).

Black-Box Testing is a software evaluation technique that assesses application functionality without analyzing the core architecture or source code. (Nguyen et al., 2023) assert that this method guarantees that specified inputs provide outputs that align with software specifications, irrespective of the system's internal operations. Testers assess only the outcomes of the evaluated functions, without requiring an understanding of the system's core workings. This approach effectively identifies functional discrepancies and ensures the program operates in accordance with user specifications (Haris & Susanto, 2021).

(Ashari et al., 2022) asserts that white box testing is a software testing methodology that emphasizes design elements and program code. It seeks to guarantee that input and output processes operate in accordance with the established requirement specifications.

This research intends to apply the K-Means Clustering algorithm to classify student reading activities at a public elementary school in Bandung, in accordance with the specified specifications. The research, moreover, creates a web-based information system proficient in digital data gathering, analysis, and dissemination of literacy activities. This project aims to streamline teacher monitoring, facilitate data-driven decision-making, and improve the efficacy of literacy programs in elementary schools.

METHOD

A. Research Methodology

This study utilizes a quantitative methodology with a descriptive framework to examine student reading activity data, including reading frequency, the quantity of books read, and categories of reading materials. The research used the K-means clustering algorithm to classify students based on their reading behavior. The data collection encompasses direct observation of the literacy recording process at SDN Arjasari Barat, interviews with the educators responsible for the literacy program, and the acquisition of documentation, including book lists, categories of reading activities, and daily logs maintained by teachers. Additionally, a literature evaluation of pertinent journals, books, and studies fortifies the theoretical framework and underpins the research analysis.

B. System Development Methodology

This study employs the Object-Oriented Analysis and Design (OOAD) methodology to systematically and modularly represent system requirements. The principal phases encompass requirements analysis, system design, implementation, and testing. In the analytical phase, the study delineates functional and non-functional criteria via observations and interviews with educators. The gathered requirements are subsequently represented using UML, including use case and object diagrams. The design phase generates class diagrams, sequence diagrams, and interface designs to serve as the foundation for the development of the web-based system. The implementation procedure entails developing literacy data-collection functionalities and applying the K-Means Clustering technique with data normalization to enhance precision. The study uses black-box testing to verify that the system meets the requirements.

RESULT AND DISCUSSION

A. Analysis of Ongoing Business Processes

The examination of existing business processes identifies and assesses all relevant procedures for the system under development. This paper delineates the functioning of legacy business processes within the Student Admissions System (PPDB) at a public elementary school in Bandung.

Table 1. Analysis of Ongoing Business Processes

| No | Entity | Activity |
|----|------------------------------|--|
| 1 | Student | <ol style="list-style-type: none"> 1. Read a book according to the reading list. 2. Report the book you have finished reading to your homeroom teacher. 3. Present your reading to the class. |
| 2 | Homeroom teacher | <ol style="list-style-type: none"> 1. Receive reading reports from students and record the number and titles of books. 2. Assess student presentations. 3. Select and group students based on the number of readings and the quality of their presentations. 4. Direct selected students to write scientific papers. 5. Submit students' scientific papers to the person in charge of Literacy. |
| 3 | Person in charge of literacy | <ol style="list-style-type: none"> 1. Receive and document scientific papers from homeroom teachers. 2. Select students with the best work. 3. Nominate the best students to perform at school or participate in literacy competitions. |

B. Problem-Solving Analysis with Role-Based Access Control (RBAC)

Table 2. Problem Solving Analysis

| Problems | Solution |
|---|---|
| The classification of students' reading achievements is inadequate, hindering teachers' ability to ascertain subsequent measures. | Employing the K-Means algorithm to categorize children according to literacy performance, specifically reading speed and content understanding. |
| No system exists that automatically assesses kids' literacy levels. | The system will compute and evaluate the centroid using literacy data, and thereafter generate clusters autonomously. |
| Educators lack a reliable benchmark for identifying pupils with high, medium, or poor talents. | The outcomes of K-Means clustering yield categories A, B, and C, which can serve as a foundation for decision-making. |
| The outcomes of K-Means clustering yield categories A, B, and C, which can serve as a foundation for decision-making. | The performance gap is determined by the disparity between a student's actual performance and the optimal standard, thereafter modified according to the cluster. |

The system uses the K-Means Clustering technique to resolve these challenges. This algorithm categorizes student literacy data—encompassing reading speed and comprehension—into three separate clusters: Grade A (high

literacy), Grade B (moderate literacy), and Grade C (poor literacy).

1) Variables Used

The two main attributes used as features in the clustering process are:

Table 3. Variables Used

| Attribute | Information |
|-----------------------------|--|
| Number of Books Read | Total Books Completed in 1 Month |
| Reading Comprehension Score | Reading Comprehension Scores From Homeroom Teachers, Scale 1 – 100 |

2) Prepared Dataset (Sample)

Table 4. Prepared Dataset (Sample)

| Student | Reading books (X) | Understanding Value (Y) |
|---------|-------------------|-------------------------|
| Tina | 12 | 90 |
| Maria | 9 | 80 |
| Iwan | 4 | 60 |
| Asep | 15 | 88 |
| Tatang | 6 | 70 |
| Lina | 3 | 50 |
| Lisda | 10 | 85 |
| Fanisa | 2 | 55 |
| Ikkal | 7 | 75 |
| Yusup | 13 | 92 |

3) Simulation and Calculation of K-Means Clustering

1. Determine the Number of Clusters (k=3)

a. Cluster 1 (Grade A) → High performance

Initial centroid: (12,90)

b. Cluster 2 (Grade B) → Mid-range performance

Initial centroid: (7,75)

c. Cluster 3 (Grade C) → Low performance

Initial centroid: (3,55)

2. Iteration 1 – Calculate the Euclidean distance from each student to the 3 centroids

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

(1)

$$\text{Distance to C1} = \sqrt{(12 - 12)^2 + (90 - 90)^2} = \sqrt{0 + 0} = \sqrt{0} \approx 0.00$$

$$\text{Distance to C2} = \sqrt{(12 - 7)^2 + (90 - 75)^2} = \sqrt{25 + 225} = \sqrt{250} \approx 15.81$$

$$\text{Distance to C3} = \sqrt{(12 - 3)^2 + (90 - 55)^2} = \sqrt{81 + 1225} = \sqrt{1306} \approx 36.14$$

b. Maria (9,80)

$$\text{Distance to C1} = \sqrt{(9 - 12)^2 + (80 - 90)^2} = \sqrt{9 + 100} = \sqrt{109} \approx 10.44$$

$$\text{Distance to C2} = \sqrt{(9 - 7)^2 + (80 - 75)^2} = \sqrt{4 + 25} = \sqrt{29} \approx 5.39$$

$$\text{Distance to C3} = \sqrt{(9 - 3)^2 + (80 - 55)^2} = \sqrt{36 + 625} = \sqrt{661} \approx 25.71$$

c. Iwan (4,60)

a. Tina (12,90)

- Distance to C1 = $\sqrt{(4 - 12)^2 + (60 - 90)^2} = \sqrt{64 + 900} = \sqrt{964} \approx 31.05$
- Distance to C2 = $\sqrt{(4 - 7)^2 + (60 - 75)^2} = \sqrt{9 + 225} = \sqrt{234} \approx 15.30$
- Distance to C3 = $\sqrt{(4 - 3)^2 + (60 - 55)^2} = \sqrt{1 + 25} = \sqrt{26} \approx 5.10$
- d. Asep (15,88)
- Distance to C1 = $\sqrt{(15 - 12)^2 + (88 - 90)^2} = \sqrt{9 + 4} = \sqrt{13} \approx 3.61$
- Distance to C2 = $\sqrt{(15 - 7)^2 + (88 - 75)^2} = \sqrt{64 + 169} = \sqrt{233} \approx 15.26$
- Distance to C3 = $\sqrt{(15 - 3)^2 + (88 - 55)^2} = \sqrt{144 + 1089} = \sqrt{1233} \approx 35.11$
- e. Tatang (6,70)
- Distance to C1 = $\sqrt{(6 - 12)^2 + (70 - 90)^2} = \sqrt{36 + 400} = \sqrt{436} \approx 20.88$
- Distance to C2 = $\sqrt{(6 - 7)^2 + (70 - 75)^2} = \sqrt{1 + 25} = \sqrt{26} \approx 5.10$
- Distance to C3 = $\sqrt{(6 - 3)^2 + (70 - 55)^2} = \sqrt{9 + 225} = \sqrt{234} \approx 15.30$
- f. Lina (3,50)
- Distance to C1 = $\sqrt{(3 - 12)^2 + (50 - 90)^2} = \sqrt{81 + 1600} = \sqrt{1681} \approx 41.00$
- Distance to C2 = $\sqrt{(3 - 7)^2 + (50 - 75)^2} = \sqrt{16 + 625} = \sqrt{641} \approx 25.32$
- Distance to C3 = $\sqrt{(3 - 3)^2 + (50 - 55)^2} = \sqrt{0 + 25} = \sqrt{25} \approx 5.00$
- g. Lisda (10,85)
- Distance to C1 = $\sqrt{(10 - 12)^2 + (85 - 90)^2} = \sqrt{4 + 25} = \sqrt{29} \approx 5.39$
- Distance to C2 = $\sqrt{(10 - 7)^2 + (85 - 75)^2} = \sqrt{9 + 100} = \sqrt{109} \approx 10.44$
- Distance to C3 = $\sqrt{(10 - 3)^2 + (85 - 55)^2} = \sqrt{49 + 900} = \sqrt{949} \approx 30.81$
- h. Fanisa (2,55)
- Distance to C1 = $\sqrt{(2 - 12)^2 + (55 - 90)^2} = \sqrt{100 + 1225} = \sqrt{1325} \approx 36.40$
- Distance to C2 = $\sqrt{(2 - 7)^2 + (55 - 75)^2} = \sqrt{25 + 400} = \sqrt{425} \approx 20.62$
- Distance to C3 = $\sqrt{(2 - 3)^2 + (55 - 55)^2} = \sqrt{1 + 0} = \sqrt{1} \approx 1.00$
- i. Ikkal (7,75)
- Distance to C1 = $\sqrt{(7 - 12)^2 + (75 - 90)^2} = \sqrt{25 + 225} = \sqrt{250} \approx 15.81$
- Distance to C2 = $\sqrt{(7 - 7)^2 + (75 - 75)^2} = \sqrt{0 + 0} = \sqrt{0} \approx 0.00$
- Distance to C3 = $\sqrt{(7 - 3)^2 + (75 - 55)^2} = \sqrt{16 + 400} = \sqrt{416} \approx 20.40$
- j. Yusup (13,92)
- Distance to C1 = $\sqrt{(13 - 12)^2 + (92 - 90)^2} = \sqrt{1 + 4} = \sqrt{5} \approx 2.24$
- Distance to C2 = $\sqrt{(13 - 7)^2 + (92 - 75)^2} = \sqrt{36 + 289} = \sqrt{325} \approx 18.03$

$$\begin{aligned} \text{Distance to C3} &= \sqrt{(13 - 3)^2 +} \\ & (92 - 55)^2 = \sqrt{100 + 1369} = \\ & \sqrt{1469} \approx 38.33 \end{aligned}$$

Table 5. Calculation of Eculidean Distance and Clustering

| Student | (x, y) | C1 (A) | C2 (B) | C3 (C) | Cluster |
|---------|--------|--------|--------|--------|---------|
| Tina | 12, 90 | 0.00 | 15.81 | 36.14 | C1 |
| Maria | 9, 80 | 10.44 | 5.39 | 25.71 | C2 |
| Iwan | 4, 60 | 31.05 | 15.30 | 5.10 | C3 |
| Asep | 15, 88 | 3.61 | 15.26 | 35.11 | C1 |
| Tatang | 6, 70 | 20.88 | 5.10 | 15.30 | C2 |
| Lina | 3, 50 | 41.00 | 25.32 | 5.00 | C3 |
| Lisda | 10, 85 | 5.39 | 10.44 | 30.81 | C1 |
| Fanisa | 2, 55 | 36.40 | 20.62 | 1.00 | C3 |
| Ikbal | 7, 75 | 15.81 | 0.00 | 20.40 | C2 |
| Yusup | 13, 92 | 2.24 | 18.03 | 38.33 | C1 |

Then update the centroid of each cluster, with the following formula:

$$\begin{aligned} X_{new} &= \frac{\sum_{i=1}^n X_i}{n}, Y_{new} \\ &= \frac{\sum_{i=1}^n Y_i}{n} \end{aligned}$$

a. Cluster 1 (C1)

Tina (12, 90), Asep (15, 88),
 Lisda (10,85), Yusup (13, 92)

$$\begin{aligned} X_{C1} &= \frac{12 + 15 + 10 + 13}{4} \\ &= \frac{50}{4} = 12.5 \\ Y_{C1} &= \frac{90 + 88 + 85 + 92}{4} \\ &= \frac{355}{4} = 88.75 \end{aligned}$$

b. Cluster 2 (C2)

Ikbal (7, 75), Maria (9, 80),
 Tatang (6,70)

$$\begin{aligned} X_{C2} &= \frac{7 + 9 + 6}{3} = \frac{22}{3} \\ &\approx 7.33 \end{aligned}$$

$$\begin{aligned} Y_{C2} &= \frac{75 + 80 + 70}{3} \\ &= \frac{225}{3} \\ &= 75.0 \end{aligned}$$

c. Cluster 3 (C3)

Iwan (4, 60), Lina (3, 50),
 Fanisa (2, 55)

$$\begin{aligned} X_{C3} &= \frac{4 + 3 + 2}{3} = \frac{9}{3} \\ &= 3.0 \end{aligned}$$

$$YC3 = \frac{60 + 50 + 55}{3}$$

$$= \frac{165}{3}$$

$$= 55.0$$

Centroid C3 (new) = (3.0,
55.0)

3. Iteration 2 – Calculate the Euclidean distance from each student to the 3 new centroids,

update the centroid values for the 1st iteration

- a. C1 : (12.5, 88.75)
 b. C2 : (7.33, 75.0)
 c. C3 : (3.0, 55.0)

Subsequently, compute the Euclidean distance from each pupil to the three newly established centroids. The subsequent table displays the outcomes.

Table 6. Results of Euclidean Distance and Clustering Calculations

| Student | (x, y) | Distance to C1 (12.5, 88.75) | Distance to C2 (7.33, 75.0) | Distance to C3 (3.0, 55.0) | Cluster |
|---------|----------|------------------------------|-----------------------------|----------------------------|---------|
| Tina | (12, 90) | 1.35 | 15.71 | 36.14 | C1 |
| Maria | (9, 80) | 9.42 | 5.27 | 25.71 | C2 |
| Iwan | (4, 60) | 29.96 | 15.37 | 5.10 | C3 |
| Asep | (15, 88) | 2.61 | 15.09 | 35.11 | C1 |
| Tatang | (6, 70) | 19.84 | 5.17 | 15.30 | C2 |
| Lina | (3, 50) | 40.05 | 25.37 | 5.00 | C3 |
| Lisda | (10, 85) | 4.51 | 10.35 | 30.81 | C1 |
| Fanisa | (2, 55) | 35.33 | 20.70 | 1.00 | C3 |
| Ikbal | (7, 75) | 14.81 | 0.33 | 20.40 | C2 |
| Yusup | (12, 90) | 1.35 | 15.71 | 36.14 | C1 |

The outcomes of the second iteration indicate that the clustering has not altered from the prior stage. The K-Means method terminates at the second iteration, achieving ideal grouping (convergence), as no individuals transitioned across clusters.

4. Conclusion of K-Mean Analysis
 After two iterations, the clustering of student data based on the two variables (\$x\$, \$y\$) reached convergence, as no individuals transitioned between groups in the second iteration.

Table 7. Final Clustering Results

| Cluster 1 (C1) | Cluster 2 (C2) | Cluster 3 (C3) |
|----------------|----------------|----------------|
| Tina | Maria | Iwan |
| Asep | Tatang | Lina |
| Lisda | Ikbal | Fanisa |
| Yusup | | |

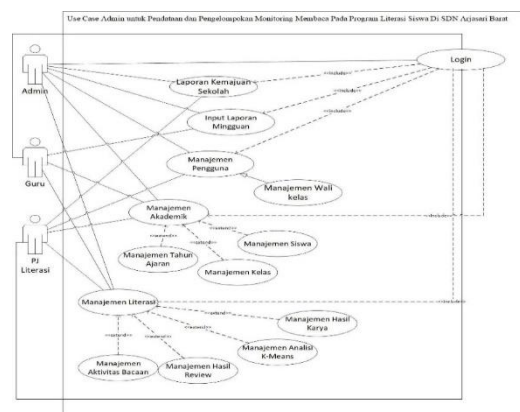
Consequently, the study finds that the K-Means method effectively categorizes 10 students into 3 clusters based on the proximity of coordinate values (x , y), which reflect two distinct factors. The procedure entails two iterations until the clustering outcomes stabilize (convergence). Each cluster exhibits unique attributes determined by its corresponding center (centroid):

- Cluster 1 (C1): Members with high scores on both variables (x and y).
- Cluster 2 (C2): Members with moderate values.
- Cluster 3 (C3): Members with low scores.

The K-Means method is useful for clustering data based on similarity, providing an objective, quantitative basis for segmentation.

C. System Design

The proposed Use Cases comprehensively delineate the actions within the system. This diagram elucidates the modeled processes by compiling use cases and actors, depicting their interrelationships within the use case diagram. The suggested Use Cases for the 'Application of the K-Means Clustering Algorithm for the Collection and Classification of Reading Monitoring Data in the Student Literacy Program at one of the public elementary schools in Bandung are as follows:

**Figure 1. Use Case Diagram Design**

D. Use Case Diagram Scenario

A scenario denotes the series of actions that delineate the interaction between the user and the system. Each scenario delineates a

particular series of events. An individual, an alternative system, hardware, or a temporal trigger initiates each sequence.

Table 8. Login Use Case Scenario Table

| | | |
|--------------------|---|-----------------------------|
| Use Case | Login | |
| Actor | Admin, Homeroom Teacher, Literacy Officer | |
| Initial Conditions | Displaying the login page | |
| Final Condition | Displaying the dashboard page | |
| Main Scenario | | |
| | Actor Action | System Action/Response |
| 1 | Opening the system page | 2 Displaying the login page |
| 3 | Fill out the login form | |
| 4 | Click the login button | 5 Login data validation |

Table 9. K-Means Analysis Use Case Scenario

| | | |
|--------------------|---|--|
| Use Case | Managing K-Means Analysis Data | |
| Actor | Admin, Homeroom Teacher, Literacy Officer | |
| Initial Conditions | Displaying K-Means Analysis Data | |
| Final Condition | Viewing the K-Means Analysis results data | |
| Main Scenario | | |
| | Actor Action | System Action/Response |
| 1 | Selecting the K-Means Analysis Menu | |
| 2 | Click the Run Analysis button | 3 Displaying Assessment Data and Clusters Student Literacy from K-Means Analysis Results |

E. Activity Diagram

1) Activity Diagram Login

The Login Activity Diagram depicts the system's response to login attempts

made by all actors, including the Admin, Teachers, and Literacy Coordinators.

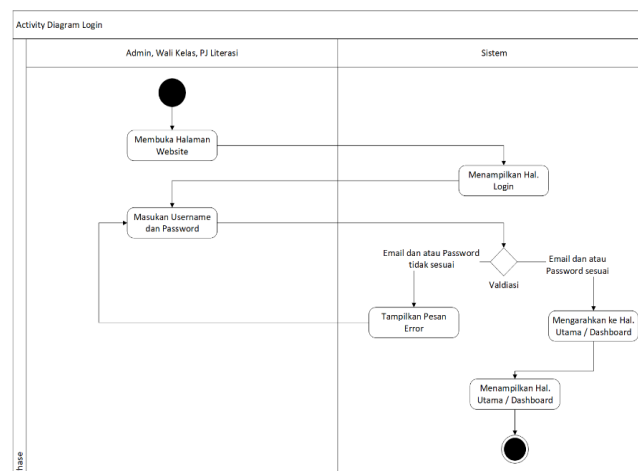


Figure 2. Activity Diagram Login

2) Activity Diagram Analisa K-Means

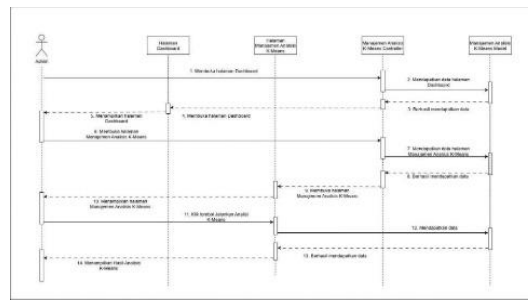


Figure 5. Sequence Diagram K-Means Analysis

G. Class Diagram

The Class Diagram depicts the interrelationships among classes in the proposed system and highlights their collaboration to achieve the designated

objectives. The subsequent diagram illustrates the Class Diagram for the 'Reading Monitoring Data Collection and Classification System for the Student Literacy Program'.

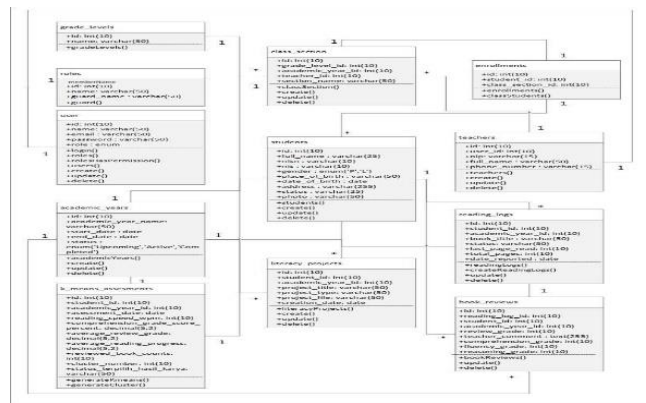


Figure 6. Class Diagram

H. Database Implementation

| # | Name | Type | Collation | Attributes | Null | Default | Comments | Extra |
|---|-------------------|--------------|--------------------|------------|------|---------|----------|----------------|
| 1 | id | bigint | | UNSIGNED | No | None | | AUTO_INCREMENT |
| 2 | name | varchar(255) | utf8mb4_unicode_ci | | No | None | | |
| 3 | email | varchar(255) | utf8mb4_unicode_ci | | No | None | | |
| 4 | email_verified_at | timestamp | | | Yes | NULL | | |
| 5 | password | varchar(255) | utf8mb4_unicode_ci | | No | None | | |
| 6 | remember_token | varchar(100) | utf8mb4_unicode_ci | | Yes | NULL | | |

Figure 7. Implementation of the Users Table

| # | Name | Type | Collation | Attributes | Null | Default | Comments | Extra |
|----|-----------------------------|--------------|--------------------|------------|------|---------|------------------------------------|----------------|
| 1 | id | bigint | | UNSIGNED | No | None | | AUTO_INCREMENT |
| 2 | student_id | bigint | | UNSIGNED | No | None | | |
| 3 | assessment_year_at | timestamp | | UNSIGNED | No | None | | |
| 4 | assessment_date | date | | | No | None | | |
| 5 | reading_speed_wpm | int | | | Yes | NULL | Kecapatan Membaca (Kata per Menit) | |
| 6 | comprehension_score_percent | decimal(5,2) | | | Yes | NULL | Dekor Pemahaman (%) | |
| 7 | average_review_grade | decimal(5,2) | | | Yes | NULL | Rata-rata Nilai Review | |
| 8 | average_reading_progress | decimal(5,2) | | | Yes | NULL | Progres Rata-rata (%) | |
| 9 | reviewed_books_count | int | | | Yes | NULL | Jumlah Buku yang Direview | |
| 10 | cluster_number | int | | | Yes | NULL | Nomor Cluster Hasil K-Means | |
| 11 | status_terpilih_hasil_karya | varchar(255) | utf8mb4_unicode_ci | | No | None | Tidak Terpilih | |

Figure 8. Implementation of K-Means Analysis Table

I. System Implementation

The System Implementation showcases the accordance with the designs outlined in the program's interface pages, created in preceding chapter.

Figure 9. Implementation of the Login System

| Nama Siswa | Judul Buku | Status | Hal. Terakhir | Total Hal. |
|------------------------------|------------|--------|---------------|------------|
| Dafinun Husni-uddin | Qui Alex C | Sudah | 121 | 194 |
| Gaflin Ikmawati | | Sedang | | |
| Gilisa Karyani Prastuti S.H. | Imawati Ta | Sedang | 168 | 201 |
| Jasmani Figenom S.H. | Darwin E | Sedang | 47 | 186 |

Figure 10. Implementation of the Weekly Reading Report System

| Nama Kelas | Jml Siswa | Rata-rata Nilai | Cluster 1 (Terpilih) | Cluster 2 | Cluster 3 |
|-------------|-----------|-----------------|----------------------|-----------|-----------|
| Kelas 1 - A | 15 | 0.00 | 0 | 0 | 0 |
| Kelas 1 - B | 15 | 0.00 | 0 | 0 | 0 |
| Kelas 2 - B | 15 | 0.00 | 0 | 0 | 0 |

Figure 11. Implementation of the School Progress System

Wiarsa¹, Emalia², Sambani³, Raharjo⁴,
 Application Of The K-Means Clustering Algorithm For Data Collection And Grouping Of Reading
 Monitoring In The Literacy Program In One Of The Public Elementary Schools In Bandung

Siswa

| Foto | Nama Lengkap | NISN | Jenis Kelamin | Status |
|------|------------------------------|------------|---------------|--------|
| | Shania Genta Widiastuti S.Pd | 2639367861 | L | Aktif |
| | Tiara Farah Hasanah | 7153288050 | P | Aktif |
| | Jamalia Pratiwi S.Gz | 8349748024 | P | Aktif |
| | Salman Prasasta | 5295209956 | L | Aktif |
| | Cinthia Zulaika | 3106452611 | L | Aktif |

Figure 12. Implementation of Student Data Management System

Catatan Membaca

| Nama Siswa | Judul Buku | Progres Membaca | Tanggal Laporan |
|--------------------|----------------------------------|-----------------|-----------------|
| Baktiadi Nugroho | Ipsium Laborum Asperiores | 51% | Jul 29, 2025 |
| Marsito Budiyo | Est Autem Incidunt | 100% | Jul 29, 2025 |
| Hasna Namaga M.TI. | Cum Ea Id | 15% | Jul 29, 2025 |
| Baktiadi Nugroho | Et Accusantium Illum | 100% | Jul 29, 2025 |
| Cinthia Zulaika | Modi Deserunt Nihil | 100% | Jul 29, 2025 |
| Hasna Namaga M.TI. | Accusantium Suscipit Perferendis | 5% | Jul 29, 2025 |
| Sidiq Nainggolan | Magni Totam Qui | 100% | Jul 29, 2025 |

Figure 13. Implementation of the Reading Notes System

Hasil Analisis K-Means
 Analisis Terakhir: 29 Juli 2025

Kelas 1 - A

| Cluster | Total Siswa | Terpilih |
|-----------|-------------|---|
| Cluster 1 | 8 siswa | Lala Sabrina Aryani, Lega Hidayat, Jasmani Firgantoro S.H., Jasmin Aryani, Mahfud Suwarno S.E., Zahra Agnes Nurdiantyanti, Umi Mayasari S.Pd, Galih Iswahyudi |
| Cluster 2 | 5 siswa | Viman Situmorang, Pranata Pranowo, Maryanto Haryanto, Nabila Aisyah Zulaika, Wardi Mujur, Dongoran M.Ak |

Kelas 1 - B

| Cluster | Total Siswa | Terpilih |
|-----------|-------------|---|
| Cluster 1 | 8 siswa | Natalia Yani Zulaika S.Psi, Almira Oliva Pertiwi, Saka Cahyono Irawan, Sarah Salimah Hartati, Maimunah Nuraini, Shania Genta Widiastuti S.Pd, Salmah Hamima Susanti, Jamalia Pratiwi S.Gz |
| Cluster 2 | 6 siswa | Bala Hutasoit, Ina Haryanti, Samiah Sudiat, Gambiang Gunarto S.Sos, Jessica Wijayanti, Salman Prasasta |

Figure 14. Implementation of the K-Means Analysis System

- J. System Testing Administrator role. This section delineates the test outcomes for the Administrator pages:
- 1) Black Box Testing System Testing entails assessments performed by the user in the

Tabel 10. Black Box Testing

| No | Tested Functions | Testing Scenario | Expected results | Test Results |
|----|---------------------------|--|--|--------------|
| 1 | Login | Leave all login data blank or fill in one of the login data (Username & password), then click the "LOGIN" button. | The system will reject Login access and display the message/notification "The Username field is required" | Succeed |
| | | Entering the condition that one of the data is correct and/or both login data are incorrect, then clicking the "LOGIN" button | The system will deny login access and display the message "The Username field is required" | Succeed |
| | | Enter the correct login data and then click the "LOGIN" button. | The system accepts the login access and then immediately displays the Administrator Main page. | Succeed |
| 2 | Create Academic Year Data | Click the create button on the Academic Year Data, then fill in the data according to the form provided, then click the "Save" button. | The system will receive a request to "create data", then the data will be saved into the database, "Data Saved Successfully", and the system will redirect to the Academic Year Data page. | Succeed |
| 3 | Create User Data | Click the create button on User Data, then fill in the data according to the form provided, then click the "Save" button. | The system will receive a request to "create data", then the data will be saved into the database, "Data Saved Successfully", and the system will redirect to the User Data page. | Succeed |
| 4 | Create Student Work Data | Click the create button on the Student Work Result Data, then fill in the data according to the form provided, then click the "Save" button. | The system will receive a request to "create data", then the data will be saved in the database, "Data Saved Successfully", and the system will redirect to the Student Work Data page. | Succeed |
| 5 | Create Book Review Data | Click the create button on the Book Review Data, then fill in the data according to the form provided, then click the "Save" button. | Sistem akan menerima akses permintaan "create data", then the data will be saved into the database, 'Data Saved Successfully', and the system will redirect to the Book Review Data page | Succeed |

| No | Tested Functions | Testing Scenario | Expected results | Test Results |
|----|-------------------------------|--|--|--------------|
| 6 | Run K-Means Analysis of Books | Clicking the 'Run K-Means Analysis' button | The system will display the results of the K-Means Assessment Data Analysis. | Succeed |

2) White Box Testing

System testing serves to demonstrate the program's performance and to verify its compliance with the specified requirements and the established design. This system uses White-Box

Testing as its assessment methodology. This White Box approach notably utilizes the Cyclomatic Complexity metric. This section outlines the program testing conducted via the White Box method:

Table 11. White Box Testing Add Academic Year

| Code Implementation | Flowchart and Flowgrah Add Academic Year |
|--|--|
| <pre> <?php namespace App\Filament\Resources\AcademicYearResource\Pages; use App\Filament\Resources\AcademicYearResource; use Filament\Actions; use Filament\Resources\Pages\CreateRecord; class CreateAcademicYear extends CreateRecord { protected static string \$resource = AcademicYearResource::class; } </pre> | |
| Independent Path | Cyclomatic Complexity |
| <p>1. 1-2-3-4-5-6-7-8-9 2. 1-2-3-4-5-6-7-5-6-8-9</p> | <p>Jumlah region = 2 $CC = E - N + 2$ $= 9 - 9 + 2$ $= 2$ $V(G) = P + 1$ $= 1 + 1 = 2$</p> |

CONCLUSION

System testing serves to demonstrate the program's performance and to verify its compliance with the specified requirements and the established design. This system uses White-Box Testing as its assessment methodology. This White Box approach notably utilizes the Cyclomatic Complexity metric. This section outlines the program testing conducted via the White Box method:

REFERENCES

- Abbas, S. A., Aslam, A., Rehman, A. U., Abbasi, W. A., Arif, S., & Kazmi, S. Z. H. (2020). K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir. *IEEE Access*, 8, 151847–151855. <https://doi.org/10.1109/ACCESS.2020.3014021>
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*, 9(8), 1295. <https://doi.org/10.3390/electronics9081295>
- Annas, M., & Wahab, S. N. (2023). Data mining methods: K-means clustering algorithms. *International Journal of Cyber and IT Service Management (IJCITSM)*, 30(1), 40-47.
- Ashabi, A., Sahibuddin, S. Bin, & Salkhordeh Haghghi, M. (2020). The Systematic Review of K-Means Clustering Algorithm. 2020 The 9th International Conference on Networks, Communication and Computing, 13–18. <https://doi.org/10.1145/3447654.3447657>
- Ashari, I. F., Banjarnahor, R., Farida, D. R., Aisyah, S. P., Dewi, A. P., & Humaya, N. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*, 6(1), 07–15. <https://doi.org/10.30871/jaic.v6i1.3485>
- Chong, B. (2021). K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science*, 4(5). <https://doi.org/10.25236/AJCIS.2021.040506>
- Ikotun, A. M., Almutari, M. S., & Ezugwu, A. E. (2021). K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Applied Sciences*, 11(23), 11246. <https://doi.org/10.3390/app112311246>
- Nguyen, T.-H. T., Dinh, D.-T., Sriboonchitta, S., & Huynh, V.-N. (2023). A method for k-means-like clustering of categorical data. *Journal of Ambient Intelligence and Humanized Computing*, 14(11), 15011–15021. <https://doi.org/10.1007/s12652-019-01445-5>
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Umargono, E., Suseno, J. E., & Vincensius Gunawan, S. K. (2020). K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. *Proceedings of the 2nd International*

*Wiarsa¹, Emalia², Sambani³, Raharjo⁴,
Application Of The K-Means Clustering Algorithm For Data Collection And Grouping Of Reading
Monitoring In The Literacy Program In One Of The Public Elementary Schools In Bandung*

Seminar on Science and Technology

(ISSTEC 2019).

<https://doi.org/10.2991/assehr.k.201010.0>

19
