Infotmatics Management, Engineering, and Information System Journal, Vol. 4, No. 1, June 2026, 53-68
E ISSN : 3025-5384, DOI : https://doi.org/10.56447/imeisj

Utilization Of The K-Means Clustering Method To Analyze The Level Of Students Memorization Progress At One Of The Al Educational Institutions In Bandung City

Pitria Ulan Sari[1], Dadi Rosadi[2], Ade Supriatna[3], Asep Suherman[4]
STMIK Mardira Indonesia, Bandung [1,2,3,4]
Email: pitriaulan@gmail.com[1], dadi.rosadi@stmik-mi.ac.id[2], ade.supriatna@stmik-mi.ac.id[3], asep.suherman@stmik-mi.ac.id[4]

Abstract

Evaluation of students' memorization abilities at an Al-Quran educational institution in Bandung was previously carried out manually, making it difficult to monitor memorization progress comprehensively and making data management inefficient. This study aims to develop a web-based information system that employs the K-Means Clustering technique to classify pupils based on their memorization proficiency. The system is designed to aid educators in categorizing students into distinct groups, including Fast Memorization Students, Normal Progress Students, Students Needing Guidance, Exceptional Memorization Students, Consistent Students, and Students Needing Motivation, to facilitate more targeted coaching strategies. The system development methodology employs an object-oriented software engineering technique, utilizing a Model-View-Controller (MVC) architecture and incorporating the Laravel framework and a MySQL database. The technology enables instructors to enter the number of memorized verses. It automatically analyzes this information utilizing the K-Means method to generate cluster divisions based on similarities in memorization values. The clustering method is executed iteratively until ideal centroids are attained. Testing is performed using Black-Box Testing techniques to verify that all system functionalities comply with user requirements. The test results demonstrate that the system operates efficiently, encompassing data input, processing via the K-Means method, and presenting outcomes in a clustered format. This technique aims to assist educators in assessing student retention more effectively and facilitate a more focused coaching approach.

Keywords : *K-Means Clustering, Student Memorization, Information System*

INTRODUCTION

Quranic educational institutes are beneficial in enhancing pupils' memorizing abilities. Nonetheless, the manual review process often yields subjective, inconsistent assessments. Consequently, accurately tracking students' memorization development is difficult, and the coaching tactics offered do not fully align with each student's needs. One such institution encounters comparable issues, as students' memorization capacities differ significantly yet remain uncharted systematically. Educators face challenges in recognizing students who are undergoing rapid advancement, stagnation, or regression. Consequently, a data-driven methodology is essential for the objective analysis and categorization of memorization advancement. The K-Means Clustering method is an efficient data mining tool for categorizing data based on similarities in specific attributes. This method can yield more precise and useful student groupings by employing characteristics such as the quantity of verses remembered, recitation frequency, error rates, and attendance. The results are anticipated to provide educators with a basis for identifying more accurate and focused coaching tactics.

A. Information System

An information system comprises individuals, facilities, or technological instruments, media, procedures, and controls

designed to structure communication networks vital for users or recipients (Zhao & Zhou, 2021).

B. Data Mining

Data mining is the process of identifying patterns and meaningful information from large datasets. It includes statistical methods, machine learning, and database systems to uncover previously undisclosed hidden knowledge(Purohit et al., 2023). In this context, data mining is employed to categorize students based on their memorization progress. This method aids TPA management in decision-making and in formulating more effective learning methods.

C. K-Means Clustering

K-means is a data analysis technique used to categorize data into several groups, where data within each group share similar characteristics. In contrast, data in other groups have distinct characteristics (Khalili Rahmatiningsih et al., 2022). Clustering is an unsupervised data mining technique, meaning the features of each cluster are not predefined; instead, the clusters represent similar attributes within a group. Consequently, analogous features from a dataset are categorized based on their properties and depicted as points in a multidimensional space (Maulana Syamil et al., 2025).

K-Means Clustering is a data clustering algorithm employed to partition a dataset into multiple clusters based on the similarity of attributes. Data within a single cluster exhibit significant homogeneity, whereas data across disparate clusters show distinct variation. This method works by identifying a predetermined number of clusters (k) and then measuring the distance of each data point to its cluster center (centroid) to assign it to the nearest group. This procedure is reiterated until the centroid locations stabilize. K-Means is widely used due to its simplicity, speed, and effectiveness in discerning patterns from large, complex datasets.

D. Object-Oriented Analysis and Design (OOAD)

Object-Oriented Analysis and Design (OOAD) is a methodology for system analysis and design employing an object-oriented paradigm. An object is an entity that possesses identity, state, and behavior. The examination of an object's identity elucidates how users differentiate it from other objects and delineates its behavior through events. Design emphasizes an object's identity as determined by its recognition by other objects for accessibility and its behavior throughout operations. This enables items to exert influence on one another inside the system (Lund & Ma, 2021). Object-Oriented Analysis and Design (OOAD) comprises a set of tools and methodologies for system development that utilizes object technology to create software systems. This has become the favored method for developing contemporary information systems (Miraftabzadeh et al., 2023). Object-Oriented Analysis and Design (OOAD) is a system design methodology centered on objects, defined as things possessing identity, attributes, and behavior, according to numerous definitions from multiple sources. Object-Oriented Analysis and Design (OOAD) is employed to construct systems utilizing an object-oriented methodology, hence enhancing their structure and facilitating development.

E. Unified Modeling Language

The Unified Modeling Language (UML) is a graphical modeling language employed to visualize and delineate software system

requirements. UML is a standard language widely used in industry to specify requirements, perform analysis and design, and illustrate architecture in object-oriented programming (Chong, 2021).

F. Laravel

Laravel is a PHP framework utilized for developing web applications. A multitude of web developers use this framework for its extensive features that enhance performance and efficiency. The framework adheres to the Model-View-Controller (MVC) architecture, which delineates data from presentation into application components, including the controller, view, model, database, and migrations (Liu, 2022).

G. Python

Python is a versatile programming language widely used by system administrators and web developers to develop dynamic websites, as well as by linguists for natural language processing applications. Python, as a general-purpose programming language, can address numerical challenges. Nonetheless, when integrated with modules such as NumPy, Seaborn, Matplotlib, and Pandas, Python can adeptly manage numerical challenges and data visualization. Consequently, Python is among the most appropriate programming languages for data visualization (Hu et al., 2023).

H. Obscured Enclosure

Black box testing is a straightforward testing methodology. The process entails analyzing the input and output of each system feature across testing scenarios, including test cases, expected outcomes, actual outcomes, and testing statuses. Testers using this method do not need knowledge of the system's source code; however, they must understand the anticipated system flow. This approach is referred to as Functional Testing, as it primarily emphasizes the system's functionality (Annas & Wahab, 2023).

I. MySQL

MySQL is a SQL database management system (DBMS) that supports multi-threading and multi-user environments, with a global user base of over 6 million. MySQL AB offers MySQL as free software under the GNU General Public License (GPL), while also providing it under a commercial license for instances where GPL usage is inappropriate. MySQL is primarily based on SQL (Structured Query Language), a crucial database language that streamlines database operations, especially for data selection and insertion, thereby enabling automated data manipulation (Purohit et al., 2023).

This study aims to use the K-Means Clustering technique to categorize students based on their memorization progress. This approach aims to assess the patterns of memorizing development derived from the clustering results, offering a more objective perspective on each student's status in the coaching process. This research aims to provide educators with pertinent and valuable insights for identifying more suitable, effective, and personalized coaching tactics tailored to each student's needs.

METHOD

Methodology

This study employs a descriptive-analytical approach. Descriptive analysis is a data processing technique that examines multiple variables to interpret study subjects and provide a comprehensive presentation of the data. This study employs descriptive analysis because it effectively delineates the actual conditions of

students' memorization progress in Quranic education facilities clearly and systematically. This method enables the data acquired to be delineated, evaluated, and understood, yielding an objective depiction of pupils' memorizing progress. The descriptive-analytical method facilitates the implementation of the K-Means Clustering algorithm used in the research, as the preliminary data analysis outcomes are essential for the student grouping procedure.

Methodology for System Development

This research employs the Object-Oriented Analysis and Design (OOAD) methodology for system development. Object-Oriented Analysis and Design (OOAD) is a methodology for system analysis and design employing an object-oriented paradigm. An object is an entity that possesses identity, state, and behavior. Examining an object's identification elucidates how users differentiate it from other entities and characterizes its activity via events. Design emphasizes an object's identity as determined by its recognition by other objects for access and its operational behavior, enabling inter-object influence within the system.

This study employs the Object-Oriented Analysis and Design (OOAD) methodology, which focuses on object-based analysis and design within the system. Each object signifies a tangible entity, including students, teachers, study materials, and assessment outcomes. The identity of an object is defined by its unique qualities, whereas its behavior is illustrated through interactions with other objects during the processes of learning and memory evaluation.

This Object-Oriented Analysis and Design (OOAD) method is integrated with the K-Means Clustering algorithm used in data analysis to categorize pupils based on their memorization progress. The grouping outcomes provide educators with valuable insights to devise more efficient and focused teaching strategies. The object-oriented methodology enables the creation of a modular, readily developed system in which each component interacts according to its designated roles and duties in analyzing pupils' memorization data.

RESULT AND DISCUSSION

1. Use Case Diagram

   A use case refers to a specific task, such as system login, data creation, or data deletion. In this sense, an actor is a human entity capable of interacting with the system to execute certain tasks. The use case diagram for the memorizing system features two actors: Admin and Educator. The administration is responsible for handling student data, including adding, modifying, and deleting entries. The following is a detailed account of the business events and the manner in which users will engage with the system.

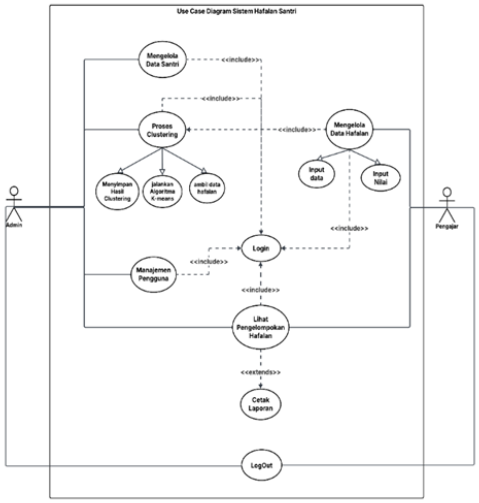


Figure 1. Use Case

2. Activity Diagram

The Activity Diagram depicts the sequence of activities within the system and the interactions between users and the system. This graphic facilitates the visualization of business processes, thereby enhancing comprehension before execution. A series of activity diagrams is presented below to elucidate the primary flow of the system, specifically:

a. Login Activity Diagram

The diagram depicts the user login process, beginning with access to the login page, followed by entering credentials (username and password), the system's verification procedure, and culminating in the user's successful entry to the main page (dashboard).

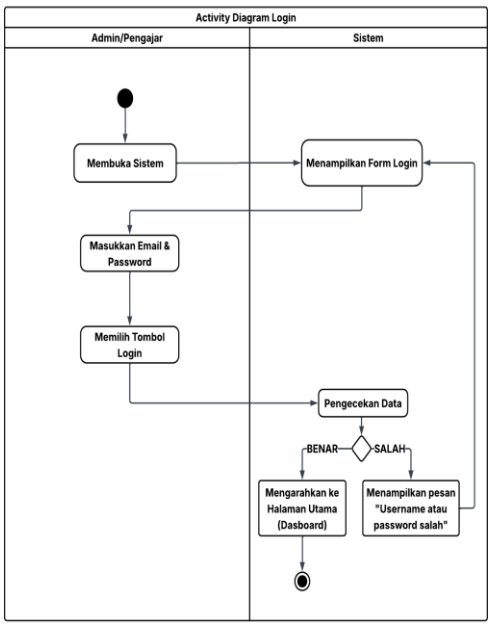


Figure 2. Login Activity Diagram

b. Clustering Process Activity Diagram

The Activity Diagram for the Clustering Process outlines the system's operations for aggregating student memorization data using the clustering methodology. The procedure commences when the administrator selects the clustering process option. The system acquires the memorization data, applies the K-Means algorithm to process and categorize the data, and subsequently presents the clustering results. The results are automatically stored in the database, and the system issues a notification confirming the successful completion of the clustering procedure.

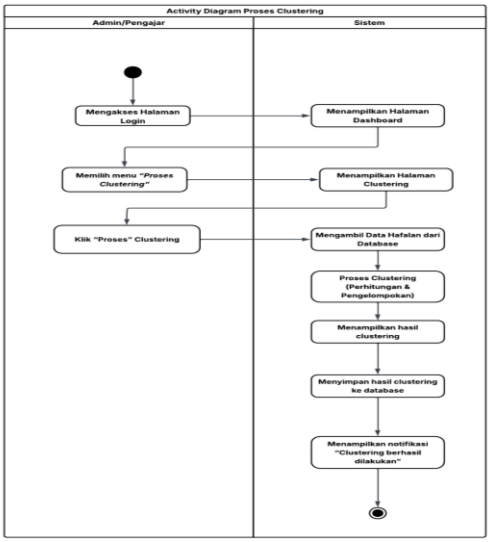


Figure 3. Activity Diagram of the Clustering Process

3. Sequence Diagram

The sequence diagram models interactions among system objects, illustrating the sequence of messages exchanged in each primary scenario. This graphic illustrates the sequential operation of each feature in accordance with the established process flow.

a. Sequence Diagram Login

This illustrates the sequence of interactions between the administrator or technician and the system during the login process, beginning with credential entry and culminating in either successful or unsuccessful verification before access to the dashboard.

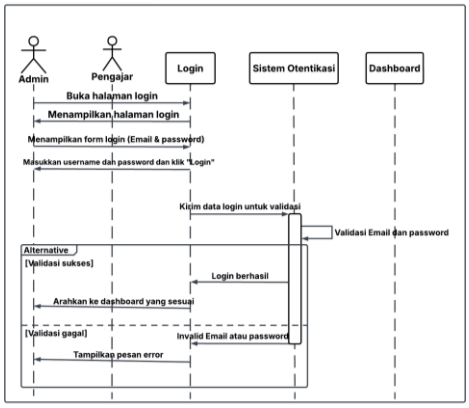


Figure 4. Login Sequence Diagram

b. Clustering Process Sequence Diagram

The graphic below depicts the flow of an action delineated in a Sequence Diagram, notably illustrating the Clustering Process action.

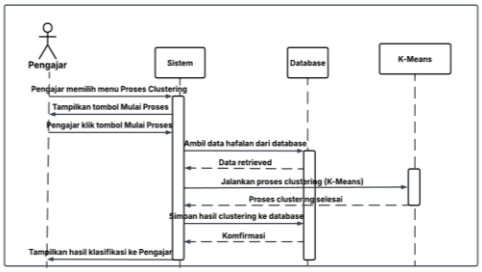


Figure 5. Sequence Diagram of the Clustering Process

4. Class Diagram

The class diagram illustrating the application of the K-Means Clustering method for analyzing students' memory progress depicts the system's architecture, including entities, characteristics, and the interrelations among them that underpin database design. This graphic is intended to guarantee that data storage functions efficiently and is structured according to the system's requirements.

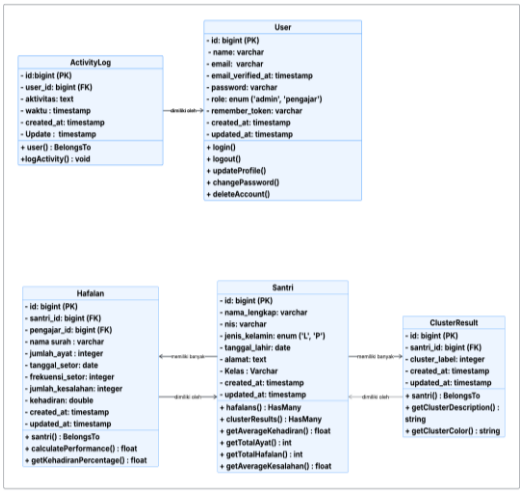


Figure 6. Class Diagram Design

5. Menu Structure

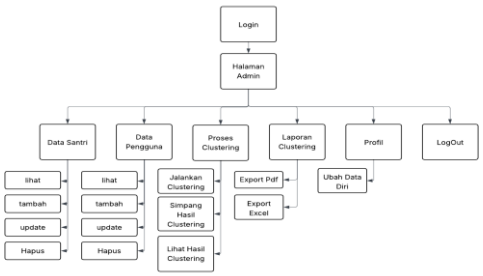


Figure 7. Admin Menu Structure Design

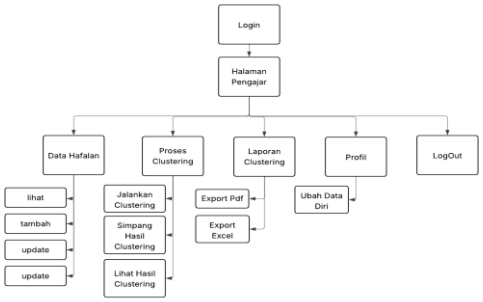


Figure 8. Design of Teacher Menu Structure

6. Database

This research involves creating a database to facilitate the analysis of students' memory progress at LPA Al-Musyawaroh. The database architecture is designed to efficiently store critical information, including student data, memorization metrics, K-Means clustering results, and memorization update operations.

Figure 9. Login Page

7. System Implementation

In the implementation phase, the interface is built according to the designs from the preceding stage to facilitate user comprehension of the system.

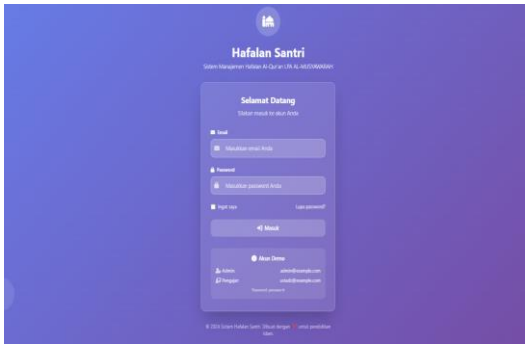


Figure 10. Login Page

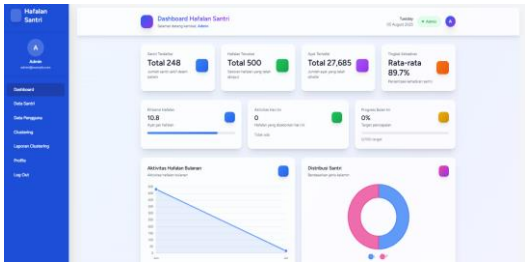


Figure 11. Dashboard Page

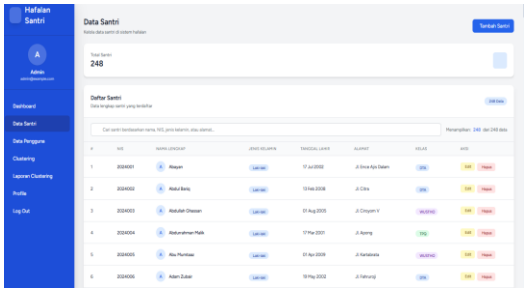


Figure 12. Student Data Page

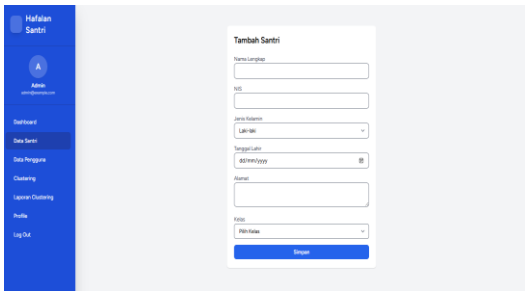


Figure 13. Add Student Page

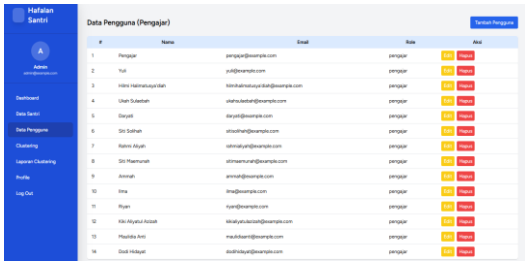


Figure 14. User Data Page

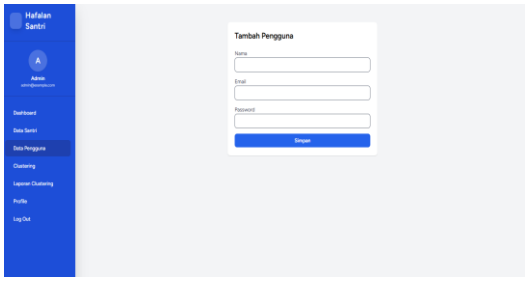


Figure 15. Add User Page

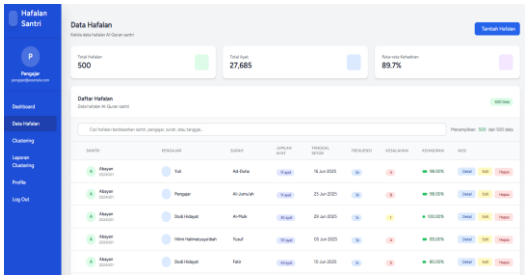


Figure 16. Memorization Data Page

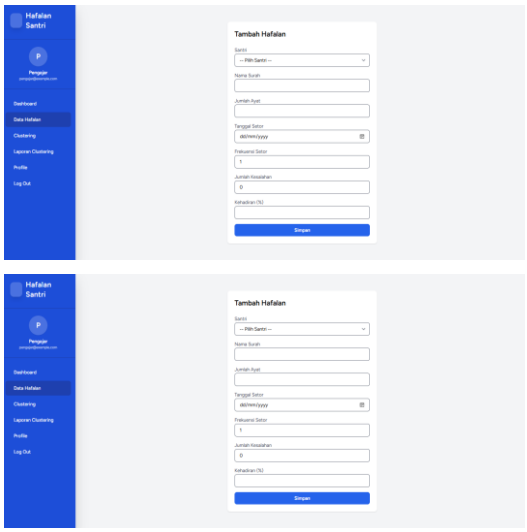


Figure 17. Memorization Data Page

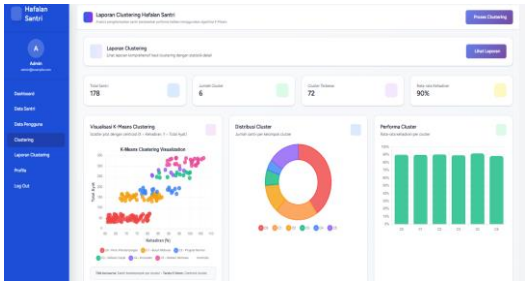


Figure 18. Clustering Page

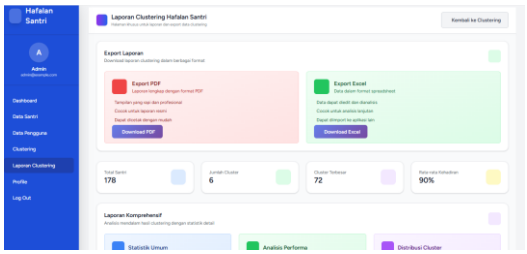
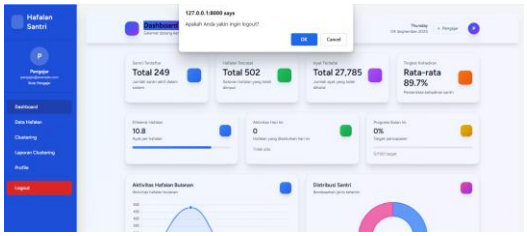


Figure 19. Clustering Report Page



Figrue 20. Logout Page

8. Implementation of the K-Means Clustering Algorithm Code

```php
// Algoritma K-Means dengan centroid yang sudah ditentukan
private function kMeansWithPredefinedCentroids($features)
{
    // Centroid yang sudah ditentukan sesuai spesifikasi
    $centroids = [
        0 => [51, 4, 9, 71],    // C0 (Perlu Pendampingan)
        1 => [148, 5, 10, 77],  // C1 (Butuh Motivasi)
        2 => [178, 1, 5, 86],   // C2 (Progres Normal)
        3 => [260, 4, 2, 94],   // C3 (Hafalan Cepat)
        4 => [270, 4, 3, 90],   // C4 (Konsisten)
        5 => [320, 5, 1, 97]    // C5 (Hafalan Istimewa)
    ];

    $clusters = [];

    // Assign setiap santri ke centroid terdekat
    foreach ($features as $santri_id => $feature) {
        $min_distance = PHP_FLOAT_MAX;
        $closest_centroid = 0;


// Vektor$santri;vfjumlah_ayat, frekuensi_setor, jumlah_kesalahan, kehadiran]
            $feature['jumlah_ayat'],
            $feature['frekuensi_setor'],
            $feature['jumlah_kesalahan'],
            $feature['kehadiran']
        ];

        // Hitung jarak Euclidean ke setiap centroid
        foreach ($centroids as $centroid_id => $centroid) {
            $distance = sqrt(
                pow($santri_vector[0] - $centroid[0], 2) +  // jumlah_ayat
                pow($santri_vector[1] - $centroid[1], 2) +  // frekuensi_setor
                pow($santri_vector[2] - $centroid[2], 2) +  // jumlah_kesalahan
                pow($santri_vector[3] - $centroid[3], 2)    // kehadiran
            );

            if ($distance < $min_distance) {
                $min_distance = $distance;
                $closest_centroid = $centroid_id;
            }
        }

        $clusters[$santri_id] = $closest_centroid;
    }

    return $clusters;
}
```

Figure 21. Implementation of the K-Means Clustering Algorithm Code

9. System Testing

   This research employs a system testing methodology centered on Black-Box Testing, executed solely based on input and output, without any insight into the program's internal architecture. This testing focuses on assessing the application's functionality by determining if each input yields the appropriate process and output consistent with the design. The objective is to ensure that the system functions in line with user requirements and expectations. The following are the system tests categorized by functionality:

Tabel 1. Pengujian Pengujian Blackbox

No	Functions tested	Testing Scenario	Expected results	Test Results
1.	Login	Empty all the login data fields or fill in one of the login data (Username & password), then click the 'Login' button.	The system will deny login access	Succeed
2.	Login	Entering the condition that one of the data is correct and/or both of the login data are incorrect, then clicking the 'Login' button	The system will deny login access and display the message “These credentials do not match our records”	Succeed

No	Functions tested	Testing Scenario	Expected results	Test Results
3.	Login	Enter the correct login data and then click the 'Login' button.	The system accepts login access and then immediately displays the main dashboard page.	Succeed
4.	Dashboard	Admin or teacher, view Registered Students, Recorded Memorization, Memorized Verses, Attendance Level, Memorization Efficiency, Today's Activity, This Month's Progress, Monthly Memorization Activity, Student Distribution, Latest Activity, Best Students, Memorization Statistics per Surah, Quick Action	The system will display the Registered Students, Recorded Memorization, Memorized Verses, Attendance Level, Memorization Efficiency, Today's Activities, This Month's Progress, Monthly Memorization Activities, Student Distribution, Latest Activities, Best Students, Memorization Statistics per Surah, Today's Quick Actions	Succeed
5.	Student Data	Admin Views student data and adds student data that has not been inputted	The system will display student data and add student data.	Succeed
6.	Edit Student Data	Admin clicks the edit button on the student data, then fills in the data according to the form provided, then clicks the "Update" button.	The system will display the form provided, then the data will be saved into the database, then after that it will display a notification message 'Data Updated Successfully, then the system will redirect to the student data page.	Succeed
7.	Delete student data	Admin clicks the delete button on the student data	The system will display a notification message "Are you sure you want to delete the student? If you click "OK" the student's data will be deleted from the database, then the system will display a notification message 'Data Successfully Deleted' then the system will redirect to the Student Data page, If you click cancel then the system will redirect to the Student Data page	Succeed
8.	User data	Admin View user data and add user data that has not been inputted	The system will display user data and add user data	Succeed
9.	Edit User Data	Admin clicks the edit button on the user data, then fills in the data according to the form provided, then clicks the "Update" button.	The system will display the form provided, then the data will be saved into the database, then after that it will display a notification message 'Data Updated Successfully, then the system will redirect to the user data page.	Succeed

No	Functions tested	Testing Scenario	Expected results	Test Results
10.	Delete User Data	Admin Clicks the delete button on user data	The system will display a notification message "Are you sure you want to delete user data? If you click "OK" the user data will be deleted from the database, then the system will display a notification message 'Data Successfully Deleted' then the system will redirect to the User Data page, If you click cancel then the system will redirect to the User Data page	Succeed
11.	Memorization data	Teachers view memorization data and add memorization data that has not been inputted.	The system will display the memorized data and add memorized data.	Succeed
12.	Edit Memorized Data	The teacher clicks the edit button on the memorization data, then fills in the data according to the form provided, then clicks the "Update" button.	The system will display the form provided, then the data will be saved into the database, then after that it will display a notification message 'Data Updated Successfully, then the system will redirect to the memorized data page.	Succeed
13.	Delete Memorized Data	Teacher clicks the delete button on the memorized data	The system will display a notification message "Are you sure you want to delete the memorized data? If you click "OK" the memorized data will be deleted from the database, then the system will display a notification message 'Data Successfully Deleted' then the system will redirect to the Memorized Data page, If you click cancel then the system will redirect to the Memorized Data page	Succeed
14.	Clustering Process	The teacher presses the “Clustering Process” button after the student's memorization data is available.	The system runs the K-Means algorithm and displays the clustering results.	Succeed
15.	View Clustering Results	Admin or Teacher opens the “Cluster Results” page and selects students	The system displays details of the classification of students according to the clustering results.	Succeed
16.	Print Clustering Report	Admin or Teacher selects the menu to print clustering results reports in PDF and EXCEL format.	The system generates clustering reports in PDF and EXCEL formats.	Succeed

No	Functions tested	Testing Scenario	Expected results	Test Results
17.	Logout	Admin or Teacher presses the "Logout" menu, then selects the confirmation Yes/Logout or Cancel	If you select Yes/Logout → the system deletes the session and displays the login page. If you select Cancel → the system cancels the logout and continues to display the dashboard.	Succeed

CONCLUSION

This study effectively created a web-based system for analyzing memorizing progress by employing the K-Means Clustering method for objective and efficient categorization. The method categorizes students' progress levels based on the quantity of memorized verses and submission timeframes, facilitating educators' comprehension of students' developmental patterns. The clustering outcomes provide a robust basis for developing more accurate coaching tactics and for enhancing the digital assessment process for memorization at a Quranic educational institution in Bandung. The depiction of clusters via an informative online interface improves monitoring efficacy and transparency in the evaluation process.

The research findings indicate that enhancing growth by incorporating supportive elements, such as direct communication between educators and students, and by broadening the analysis to encompass external influences, can render the analysis more thorough. Enhancements to clustering categories and the integration of the system with pesantren administrative platforms are essential for optimal, comprehensive deployment in Islamic educational settings. The system possesses the potential to evolve into a more precise, adaptable, and efficient memorization assessment instrument through these developmental chances.

REFERENCES

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.

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

Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis. Pattern Recognition, 139, 109404. https://doi.org/10.1016/j.patcog.2023.109404

Khalili Rahmatiningsih, A., Nilogiri, A., & Eko Wardoyo, A. (2022). Pengelompokan Hasil Evaluasi Pembelajaran Metode Hafalan Al Qur'an Tawazun Menggunakan Metode K-Means. Jurnal Indonesia Sosial Teknologi, 3(08), 879–894. https://doi.org/10.36418/jist.v3i8.480

Liu, R. (2022). Data Analysis of Educational Evaluation Using K-Means Clustering Method. Computational Intelligence and Neuroscience, 2022, 1–10. https://doi.org/10.1155/2022/3762431

Lund, B., & Ma, J. (2021). A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering. Performance Measurement and Metrics, 22(3), 161–173. https://doi.org/10.1108/PMM-05-2021-0026

Maulana Syamil, B., Susilo Yuda Irawan, A., & Mayasari, R. (2025). ALGORITMA K-MEANS UNTUK MENGUKUR TINGKAT KEPUASAN SISWA DAN GURU DALAM MEMANFAATKAN GAFE (GOOGLE APPS FOR EDUCATION). JATI (Jurnal Mahasiswa Teknik Informatika), 9(2), 2506–2515. https://doi.org/10.36040/jati.v9i2.13167

Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foiadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. IEEE Access, 11, 119596–119633. https://doi.org/10.1109/ACCESS.2023.3327640

Purohit, K., Vats, S., Saklani, R., Kukreja, V., Sharma, V., & Yadav, S. P. (2023). Improvement in K-Means Clustering for Information Retrieval. 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 1239–1245. https://doi.org/10.1109/ICESC57686.2023.10193031

Zhao, Y., & Zhou, X. (2021). K-means Clustering Algorithm and Its Improvement Research. Journal of Physics: Conference Series, 1873(1), 012074. https://doi.org/10.1088/1742-6596/1873/1/012074