

## K-Means Clustering Analysis for Mapping Alumni Career Success Profiles to Evaluate Graduate Quality at SMKN 1 Kisaran

Dian Toro Putra Wiguna <sup>1\*</sup>, Dahriansah <sup>2</sup>, Muhammad Bayu Al Dhana <sup>3</sup>

<sup>1\*,2</sup> Information Systems Study Program, Faculty of Computer Science, Universitas Royal, Asahan Regency, North Sumatra Province, Indonesia

<sup>3</sup> Mathematics Education Study Program, Faculty of Teacher Training and Education, Universitas Royal, Asahan Regency, North Sumatra Province, Indonesia

Email: wigunatoro82@gmail.com <sup>1\*</sup>, andriok818@gmail.com <sup>2</sup>, bayualdhana0222@gmail.com <sup>3</sup>

### Article history:

Received September 2, 2026

Revised October 12, 2026

Accepted October 15, 2026

### Abstract

The development of information and communication technology has encouraged the widespread use of data in various fields, including education. This study aims to apply the K-Means Clustering method to group alumni career data of SMKN 1 Kisaran based on predetermined indicators of career success. The research uses a quantitative approach with a data mining method, applying the K-Means Clustering algorithm as the main technique in the data analysis process. The findings indicate that the alumni career data were successfully grouped by converting categorical data into numerical weights for computational processing. The results show that the alumni data are divided into three main clusters. Cluster 1 represents alumni with a high level of career success, characterized by a short waiting period for employment and suitability between their field of work and competencies. Cluster 2 describes alumni with a moderate level of career success, indicated by a longer waiting period for employment or a mismatch between their field of work and competencies. Cluster 3 represents alumni with a low level of career success, dominated by alumni who have not yet obtained employment.

### Keywords:

K-Means Clustering Method; Data Mining; Career Profile Mapping; Alumni Career Success; Graduate Quality Evaluation; SMKN 1 Kisaran.

## 1. INTRODUCTION

The rapid development of information and communication technology has increased the use of data in education. Schools are now expected not only to produce capable graduates, but also to assess graduate quality in a clear, systematic, and evidence-based way (Han et al., 2022). This condition makes information systems and data mining increasingly relevant for managing alumni data, especially as the volume of records continues to grow each year and cannot be handled well through manual review alone. Vocational High Schools (SMK) have a key role in preparing graduates with practical skills that meet labor market demands. For that reason, alumni career success becomes one of the main indicators for evaluating whether the educational process has worked as intended (Ministry of Education and Culture of the Republic of Indonesia, 2021).

Vocational High Schools aim to prepare students with professional competencies, practical skills, and work attitudes that are aligned with the needs of the business and industrial world. In this regard, the success of SMK graduates should not be assessed solely from graduation rates, but also from how well alumni are able to enter the workforce after completing their studies. Important indicators include the speed of obtaining employment, the suitability between the field of work and the competencies learned at school, as well as the current employment status of alumni. These indicators provide a clearer picture of whether vocational education has truly met its intended goals. Because of this, the management and analysis of alumni career data become essential steps in supporting graduate quality evaluation and improving the quality of vocational education on an ongoing basis. Proper data processing can help schools identify patterns, assess weaknesses, and formulate better strategies for preparing future graduates (Prasetyo, 2020).

SMK Negeri 1 Kisaran, a vocational school in Asahan Regency, produces a large number of alumni every year who spread across various sectors, including employment, higher education, and entrepreneurship. To monitor their outcomes, the school has collected alumni career data through tracer studies and routine alumni data collection. However, the available data is still mostly presented in descriptive reports and simple tables. As a result, the information has not yet been processed in a way that clearly shows patterns of alumni career success. This condition makes it difficult for the school to distinguish alumni groups with high, medium, or low levels of career achievement. Without a clearer pattern analysis, the school may find it harder to use the data as a basis for evaluating graduate quality and improving vocational learning strategies. Therefore, a more structured data analysis is needed so that alumni career information can be used more effectively and provide a stronger basis for decision-making (Fatkhudin et al., 2023).

The diversity of this data also makes manual or purely descriptive analysis less effective, especially when the number of alumni records continues to grow and the characteristics of each record become increasingly varied. Relying only on simple summaries can leave important patterns unnoticed and limit the school's ability to understand alumni career distribution in a meaningful way. For that reason, a technology-based approach using clustering methods is needed so alumni data can be grouped systematically, consistently, and measurably into specific career success profiles. Through this approach, similar data can be placed in the same group, while data with different characteristics can be separated more clearly. The use of data mining techniques in analyzing alumni career data is considered capable of producing a more objective and accurate mapping of career success profiles, which can later serve as a practical basis for evaluating graduate quality and improving school decisions related to vocational education development (Devie & Anamisa, 2020).

The K-Means method is one of the most widely used clustering algorithms because it is simple, efficient, and relatively easy to implement in various data analysis tasks (Rudini, 2024). This method works by dividing data into several clusters based on the closest distance to the cluster center, or centroid, so that records with similar characteristics are placed in the same group. Through this process, the data can be organized more clearly and analyzed with greater consistency. In the case of alumni career data, the use of K-Means allows alumni to be grouped according to their level of career success in a measurable and objective way. Such grouping helps schools see differences in alumni outcomes more clearly and use the results as a basis for evaluating graduate quality. In addition, the method supports faster data processing and provides results that are easier to interpret for educational decision-making (Soufitri, 2023).

Mapping alumni career success profiles through K-Means clustering can provide useful strategic information for SMK Negeri 1 Kisaran. Through this process, the school can identify groups of alumni who are quickly absorbed into the workforce, alumni who continue their education, and alumni who still face difficulties in finding employment. Such grouping does not only show the distribution of alumni outcomes, but also helps reveal which career patterns are more dominant among graduates. The results can then be used as evaluation material for graduate quality, especially in assessing whether the school has succeeded in preparing students for real labor market demands. In addition, the clustering results can serve as a basis for formulating policies to improve educational quality, including curriculum adjustment, career guidance, and stronger links with the business and industrial world. If analyzed properly, this information can help the school make more accurate decisions and develop follow-up strategies that are grounded in alumni data rather than assumptions alone (Novitasari et al., 2023).

This study aims to apply the K-Means Clustering method as a data analysis technique for grouping alumni career records at SMKN 1 Kisaran based on predetermined career success indicators. Through this grouping process, the study is expected to produce a mapping of alumni career success profiles into several objective, clear, and measurable clusters that represent the characteristics of each alumni group more accurately. The results of this analysis are also expected to present data-based information in a form that is easier to understand and use by the school. In practice, the findings can serve as evaluation material for assessing graduate quality, identifying alumni career patterns, and supporting the development of education quality improvement strategies. In addition, the output of the clustering process can help the school design follow-up programs that are more relevant to labor market needs and better aligned with alumni career development after graduation.

Based on the description above, the author conducted a thesis research entitled "K-Means Clustering Analysis for Mapping Alumni Career Success Profiles as an Evaluation of Graduate Quality Based on Alumni Career Data at SMKN 1 Kisaran." This research was designed to examine alumni career data more systematically so that the information collected by the school can be processed into clearer and more useful results. By applying the K-Means method, the study is expected to produce a structured mapping of alumni career success profiles that can describe the distribution of graduate outcomes in a more objective way. The findings are also expected to provide a real contribution to the optimal use of alumni data, especially as a reference for evaluating graduate quality and improving school policies. In this way, SMK Negeri 1 Kisaran can strengthen its educational quality development based on empirical data rather than assumptions alone (Pratiwi & Mulyawan, 2023).

## 2. RESEARCH METHOD

The research method section is structured systematically to ensure that each stage of the research process runs in a directed, logical, and accountable manner. The process begins with identifying the research problem, followed by collecting and organizing the relevant alumni career data, and then analyzing the data using the K-Means Clustering method to group alumni based on career success characteristics. This study applies a quantitative approach supported by data mining techniques, with the K-Means algorithm as the main method used to process and classify the data objectively (Han et al., 2022; Prasetyo, 2020). Through this approach, the research is expected to produce clear and measurable results that can be used to interpret alumni career patterns. The explanation is as follows.

### 2.1. Identification of Problems

The problem identification stage was conducted to recognize various issues related to the management and utilization of alumni career data at SMKN 1 Kisaran. At this stage, observations were made to understand how alumni data were being processed and to see whether the information had been used optimally for analysis. The initial findings showed that the data were still presented in a descriptive form and had not yet been examined more deeply to reveal patterns of alumni career success. As a result, the school did not yet have a clear mapping of alumni outcomes that could support evaluation activities. Therefore, problem identification becomes an important stage to formulate the research focus, define the analytical needs, and determine the most suitable approach for helping the school conduct a more objective, structured, and data-based evaluation of graduate quality (Ministry of Education and Culture of the Republic of Indonesia, 2021).

### 2.2. Data Collection

The data collection stage was carried out by gathering information related to the careers of SMKN 1 Kisaran alumni in a structured manner. The data collected included several important aspects, such as the length of time alumni waited to get a job, the suitability between their current job field and their area of competency expertise, and their employment status after graduation. These data were obtained through tracer study activities, alumni documentation, and other relevant school records that could support the analysis process. In addition, the collected information was organized carefully so that it could be used as the main material in data mining analysis. By using complete and relevant data, the research is expected to produce more accurate results and support a better understanding of alumni career patterns (Prasetyo, 2020).

### 2.3. Data and System Analysis

The data and system analysis phase aims to understand the characteristics of the alumni career data that has been collected and determine the variables to be used in the analysis process. At this stage, data selection, data cleaning, and data format adjustment are carried out so that the data is ready to be processed using the K-Means clustering method. In addition, an analysis of the system requirements to be developed is conducted to ensure that the system is capable of managing alumni data and carrying out the data grouping process effectively (Han et al., 2022; Wakhidah, 2020).

### 2.4. System Design

The system design stage is carried out to arrange the structure, workflow, and technical framework of the system that will be developed. At this stage, the design includes data processing procedures, database structure, and the way the K-Means method will be applied in the system so that the analysis process can run properly and efficiently. In addition, the system model is designed using modeling tools such as the Unified Modeling Language (UML), which consists of use case diagrams, activity diagrams, and class diagrams. These models are used to describe how users interact with the system, how the process flows from one stage to another, and how the data are organized within the application. Through this design stage, the researcher can obtain a clearer picture of the system to be built, reduce implementation errors, and ensure that the final system is aligned with the research objectives (Rudini, 2024; Soufitri, 2023).

### 2.5. System Development

The system development stage is the realization of the system design that has been prepared in the previous phase. At this stage, a web-based application is developed to manage alumni data more effectively and to apply the K-Means method in the process of grouping the data. The application is built using the PHP programming language, supported by the CodeIgniter framework to simplify development and improve the structure of the system. Meanwhile, MySQL is used as the database to store, organize, and retrieve alumni data needed in the analysis process. Through this development stage, the system begins to take a functional form and can be tested according to its intended purpose. The development process also ensures that the application is able to support data processing activities in a more practical, efficient, and organized way, so the resulting system can be used as a reliable tool for analyzing alumni career patterns (Devie & Anamisa, 2020; Rudini, 2024).

## 2.6. System Testing

The system testing phase is conducted to ensure that the developed system can operate properly and in line with the research requirements. Testing is carried out by examining each function of the system in detail, starting from the data input stage, continuing to data processing, and ending with the clustering process using the K-Means method. Through this stage, the researcher can observe whether the system is able to process alumni data correctly and produce grouping results that are accurate and consistent with the expected analysis objectives. In addition, system testing is also intended to identify possible errors or weaknesses that may appear during the data processing process so that they can be corrected before the system is used more broadly. This stage is important because it helps ensure that the application is not only functional, but also reliable in supporting alumni data analysis and graduate quality evaluation (Fatkhudin et al., 2023; Harini & Kusumawati, 2023).

## 2.7. System Evaluation

The system evaluation stage is conducted to assess the clustering results of alumni career data produced by the system. At this stage, the results of data grouping are reviewed carefully to identify the patterns of alumni career success that appear in each cluster. The evaluation process is not only focused on the accuracy of the clustering output, but also on how useful the results are for interpreting alumni career conditions in a more structured way. Through this stage, the researcher can determine whether the system is able to present information that reflects actual alumni career characteristics. The results of this evaluation are expected to provide useful input for schools in evaluating graduate quality and understanding the relation between alumni career outcomes and school preparation. In addition, the evaluation can serve as a reference for improving educational quality and strengthening graduate readiness to face the demands of the world of work (Novitasari et al., 2023; Pratiwi & Mulyawan, 2023).

# 3. RESULTS AND DISCUSSION

## 3.1. Result

The analysis results were obtained from the processing of alumni career data at SMKN 1 Kisaran using the K-Means Clustering method. The data analysis process involved preparing alumni career variables, transforming categorical data into numerical values, and applying the clustering algorithm to classify alumni based on career success indicators. The clustering results were used to identify groups of alumni with similar career characteristics after graduation. Each cluster was then analyzed based on employment status, waiting period before obtaining a job, and the alignment between occupational fields and school competencies. The results of the grouping process describe the distribution of alumni career profiles and can be used as a reference for evaluating graduate quality at SMKN 1 Kisaran.

### 3.1.1. Analysis of the Current System

The current management of alumni career data at SMKN 1 Kisaran has not fully utilized data analysis methods to identify graduate career patterns. Alumni information is mainly stored as administrative records and processed through descriptive reports. This process makes it difficult to classify alumni based on career success indicators, including employment status, waiting time to obtain a job, and the relevance between occupational fields and competencies acquired during education. The absence of a clustering approach also causes alumni data analysis to depend mainly on manual summaries. As a result, the school does not yet have a clear classification of alumni groups based on their career conditions. The application of K-Means Clustering can assist the grouping process by classifying alumni records according to similarities in career characteristics and producing measurable cluster information. The current tracer study process begins with alumni completing questionnaires containing information related to post-graduation conditions, including employment status, further education, and feedback regarding learning experiences at school. The collected data are then managed by administrative staff and recorded into Microsoft Excel for reporting purposes. After the data compilation process is completed, a recapitulation report is prepared and submitted to the principal. The principal uses the report as supporting information for evaluating graduate conditions and planning improvements related to education quality. The current system flow is illustrated in the following figure1.

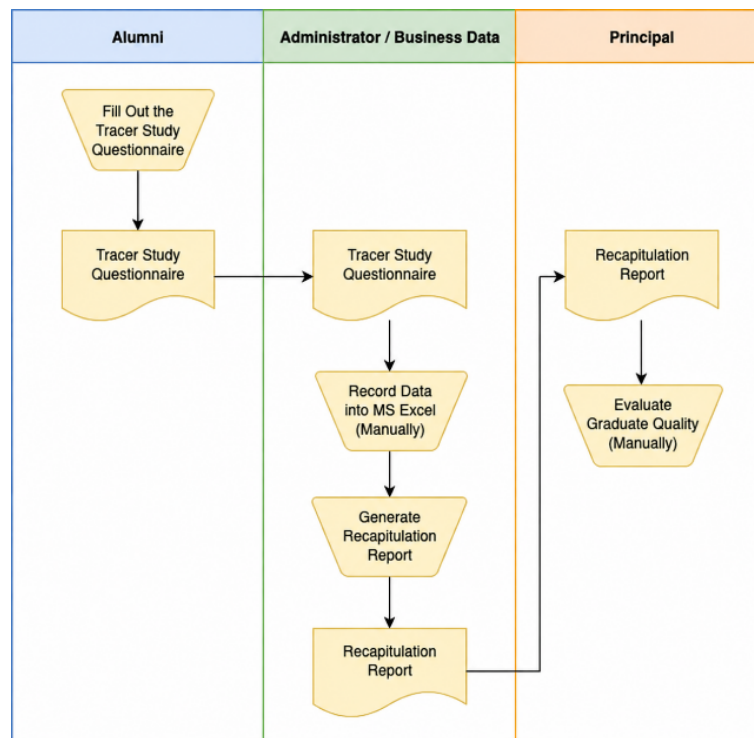


Figure1. Current System Flow

### 3.1.2. Proposed System Analysis

Based on the analysis of the current system, a system is required to process alumni career data in a more structured, systematic, and data-driven manner. The proposed system applies the K-Means Clustering method to classify alumni data based on career success indicators, including employment waiting period, alignment between occupational fields and competency expertise, and employment status. Through the clustering process, alumni data that were previously stored only as administrative records can be converted into meaningful information in the form of groups with similar characteristics. The clustering results are expected to provide an objective representation of alumni career success profiles and assist SMKN 1 Kisaran in conducting graduate quality evaluations more accurately. Furthermore, the proposed system can improve data processing efficiency, minimize errors caused by manual analysis, and generate reports that are easier to interpret as supporting information for decision-making. Based on the proposed system flow, the tracer study process is implemented through an integrated web-based system to improve alumni data management and support graduate quality evaluation activities. The process begins when alumni access the system portal using their registered accounts and complete the tracer study questionnaire based on the available data fields. After the submission process is completed, the collected information is automatically stored in the database as the main data repository. The administrator then accesses the system to manage the stored alumni data and performs the data transformation process by converting categorical attributes into numerical values using predetermined weighting criteria according to the requirements of the K-Means algorithm.

After the transformation stage is completed, the system executes the K-Means Clustering algorithm to group alumni records based on the similarity of their career characteristics. The clustering process produces several groups that represent different levels of alumni career success. The generated cluster results are stored in the database and can be accessed for further analysis. The principal can then log in to the system to view the dashboard containing tracer study information and clustering results. If required, the principal can generate and print clustering reports as supporting documents for graduate quality evaluation. The proposed information system flow is illustrated in the following figure 2.

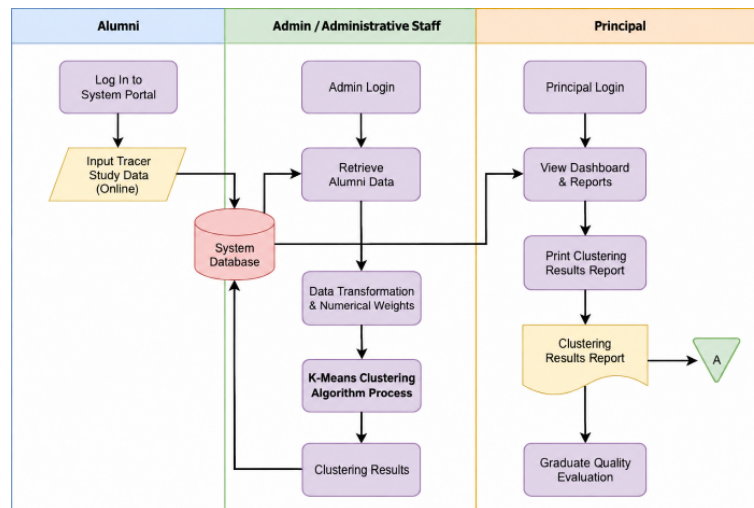


Figure 2. Proposed System Flow

### 3.1.3. Use Case Diagram

The Use Case Diagram is used to describe the interaction between actors and the proposed system, as well as to explain the functions that can be performed by each user. Based on the analysis of the current system, the proposed system is designed to overcome limitations in managing alumni career data, which was previously processed manually and did not utilize data-driven analysis methods. The diagram illustrates the main activities within the system, including the login process, alumni data management, data processing using the K-Means Clustering method, presentation of clustering results, and graduate quality evaluation reports. In addition, the Use Case Diagram defines the access rights of each actor to ensure that every system function operates according to the responsibilities and authorities of each user. The proposed Use Case Diagram design for the system is presented in the following figure 3.

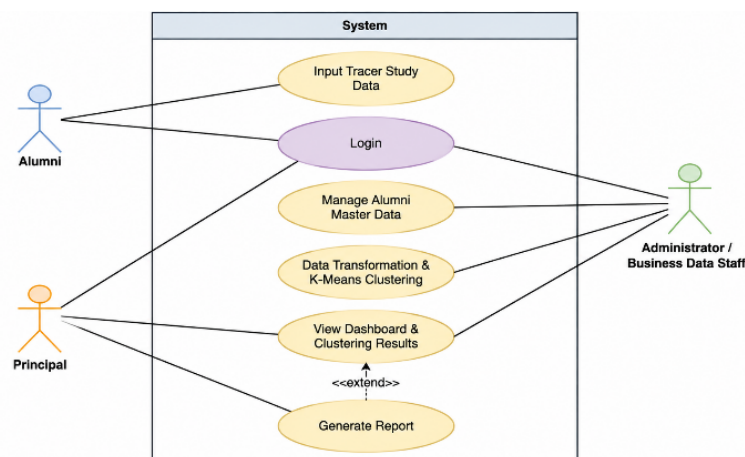


Figure 3. Use Case Diagram

### 3.1.4. Use of the K-Means Clustering Method

The K-Means Clustering method is applied in this study as a data mining technique to group alumni career data based on the similarity of their characteristics. The clustering process uses several career success indicators, including employment waiting period, alignment between occupational fields and competency expertise, employment status, and other relevant variables determined in the study. The implementation of the method begins with determining the number of clusters, followed by selecting the initial centroid values. The distance between each data point and the centroid is then calculated using the Euclidean Distance method to determine the nearest cluster. After the initial grouping process, the centroid values are updated based on the average values of data points within each cluster. The iteration continues until no changes occur in cluster membership or the centroid values reach a stable condition. The clustering results produce a structured mapping of alumni career success profiles based on their characteristics. The application of the K-Means method enables SMKN 1 Kisaran to identify different groups of alumni career outcomes and use the results as supporting information for graduate quality evaluation. Furthermore, the clustering process helps identify patterns in alumni career conditions that can be considered in developing educational improvement strategies aligned with workforce requirements. The following section explains the stages of the K-Means Clustering method.

Once all the data is in the form of a numerical matrix, the system will determine the number of clusters ( $K = 3$ ) and select three initial centroids. These starting points are randomly selected from the available data. For this first iteration of the simulation, the starting points (centroids).

Table 1. Determination of Initial Centroid (Iteration 1)

| Centroid | X1 | X2 | X3 | Information                                    |
|----------|----|----|----|------------------------------------------------|
| C1       | 5  | 0  | 4  | High Indication, Taken from A14 data value     |
| C2       | 5  | 7  | 3  | Moderate Indication, Taken from A25 data value |
| C3       | 1  | 0  | 1  | Low Indication, Taken from data value A15      |

Each alternative data is allocated to the closest cluster. The following is a description of the manual calculation for finding the distance between the first data object (A1) and the last data object (A25) to the three centroids (C1, C2, C3). After obtaining the distance results in iteration 1, the next stage is to find the minimum value between C1, C2, and C3.

Table 2. First Iteration K-Means Calculation (Minimum &amp; Cluster)

| Alternative | C1    | C2   | C3    | Minimum | Cluster |
|-------------|-------|------|-------|---------|---------|
| A1          | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A2          | 3.00  | 7.28 | 4.00  | 3.00    | 1       |
| A3          | 1.41  | 7.07 | 3.61  | 1.41    | 1       |
| A4          | 2.00  | 7.07 | 4.12  | 2.00    | 1       |
| A5          | 2.00  | 7.07 | 4.12  | 2.00    | 1       |
| A6          | 1.41  | 6.00 | 4.58  | 1.41    | 1       |
| A7          | 5.10  | 3.74 | 5.00  | 3.74    | 2       |
| A8          | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A9          | 1.00  | 7.14 | 4.24  | 1.00    | 1       |
| A10         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A11         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A12         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A13         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A14         | 0.00  | 7.07 | 5.00  | 0.00    | 1       |
| A15         | 5.00  | 8.31 | 0.00  | 0.00    | 3       |
| A16         | 12.04 | 5.00 | 12.81 | 5.00    | 2       |
| A17         | 0.00  | 7.07 | 5.00  | 0.00    | 1       |
| A18         | 2.00  | 7.07 | 4.12  | 2.00    | 1       |
| A19         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A20         | 1.41  | 7.07 | 3.61  | 1.41    | 1       |
| A21         | 0.00  | 7.07 | 5.00  | 0.00    | 1       |
| A22         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A23         | 1.00  | 7.14 | 4.24  | 1.00    | 1       |
| A24         | 1.00  | 7.00 | 4.47  | 1.00    | 1       |
| A25         | 7.07  | 0.00 | 8.31  | 0.00    | 2       |

Based on the distance between each data point and the initial centroid, each data point is placed in the cluster with the shortest distance. The first iteration's calculations show that alumni career success is grouped into three clusters. The clustering results from the first iteration are then used to calculate new centroids, which will be used in subsequent iterations until no cluster changes occur.

Table 3. New Cluster Centers of the Second Iteration

| Alternative | C1 | C2 | C3 |
|-------------|----|----|----|
| A1          | 1  |    |    |
| A2          | 1  |    |    |
| A3          | 1  |    |    |
| A4          | 1  |    |    |
| A5          | 1  |    |    |
| A6          | 1  |    |    |
| A7          |    | 1  |    |
| A8          | 1  |    |    |
| A9          | 1  |    |    |
| A10         | 1  |    |    |
| A11         | 1  |    |    |
| A12         | 1  |    |    |
| A13         | 1  |    |    |

|     |   |   |
|-----|---|---|
| A14 | 1 |   |
| A15 |   | 1 |
| A16 |   | 1 |
| A17 | 1 |   |
| A18 | 1 |   |
| A19 | 1 |   |
| A20 | 1 |   |
| A21 | 1 |   |
| A22 | 1 |   |
| A23 | 1 |   |
| A24 | 1 |   |
| A25 |   | 1 |

After the distance calculation process is carried out in the first iteration, each data point has been grouped into the cluster with the closest distance to the centroid. The next step is to calculate a new centroid for each cluster based on the average attribute values of the cluster members. The new centroid calculation is performed using the following average formula.

Table 4. New Centroid C1 (21 Data)

| Attribute                | Calculation                                              | Results |
|--------------------------|----------------------------------------------------------|---------|
| Graduate Activity Status | $(5+5+4+5+5+5+5+4+5+5+5+5+5+5+5+5+5+5+5+4+5+5+4+5) / 21$ | 4.81    |
| Waiting Period (Months)  | $(0+0+0+0+0+1+0+0+0+0+0+0+0+0+0+0+0+0+0+0+0+0+0+0) / 21$ | 0.05    |
| Level of Alignment       | $(3+1+3+2+2+3+3+4+3+3+3+3+4+4+2+3+3+4+3+4+3) / 21$       | 3.0     |

Based on the calculation results, new centroid values are obtained from the first iteration, which will be used as reference points for the next iteration process. These updated centroids represent the average characteristics of each cluster after the initial grouping stage. The recalculated centroid values are then applied in iteration 2 to determine the next clustering process. The detailed centroid results are presented in the following table 5.

Table 5. New Centroids for Iteration 2

| Centroid | Graduate Activity Status | Waiting Period (Months) | Level of Alignment |
|----------|--------------------------|-------------------------|--------------------|
| C1       | 4.81                     | 0.05                    | 3.00               |
| C2       | 4.67                     | 7.67                    | 2.33               |
| C3       | 1.00                     | 0.00                    | 1.00               |

The second iteration begins by recalculating the distance between each data point and the updated centroid values obtained from the previous iteration. The calculation process follows the same procedure as the third step, using the Euclidean Distance formula described in the previous equation. The purpose of this stage is to determine whether each data point remains in the same cluster or moves to another cluster based on the nearest centroid. The results of the distance calculation in the second iteration are presented in the following table 6.

Table 6. K-Means Calculation (Minimum &amp; Cluster)

| Alternative | C1    | C2   | C3    | Minimum | Cluster |
|-------------|-------|------|-------|---------|---------|
| A1          | 0.20  | 7.70 | 4.47  | 0.20    | 1       |
| A2          | 2.01  | 7.79 | 4.00  | 2.01    | 1       |
| A3          | 0.81  | 7.72 | 3.61  | 0.81    | 1       |
| A4          | 1.02  | 7.68 | 4.12  | 1.02    | 1       |
| A5          | 1.02  | 7.68 | 4.12  | 1.02    | 1       |
| A6          | 0.97  | 6.71 | 4.58  | 0.97    | 1       |
| A7          | 4.50  | 3.96 | 5.00  | 3.96    | 2       |
| A8          | 0.20  | 7.70 | 4.47  | 0.20    | 1       |
| A9          | 1.29  | 7.87 | 4.24  | 1.29    | 1       |
| A10         | 0.20  | 7.70 | 4.47  | 0.20    | 1       |
| A11         | 0.20  | 7.70 | 4.47  | 0.20    | 1       |
| A12         | 0.20  | 7.70 | 4.47  | 0.20    | 1       |
| A13         | 0.20  | 7.70 | 4.47  | 0.20    | 1       |
| A14         | 1.02  | 7.85 | 5.00  | 1.02    | 1       |
| A15         | 4.30  | 8.60 | 0.00  | 0.00    | 3       |
| A16         | 11.95 | 4.40 | 12.81 | 4.40    | 2       |
| A17         | 1.02  | 7.85 | 5.00  | 1.02    | 1       |
| A18         | 1.02  | 7.68 | 4.12  | 1.02    | 1       |
| A19         | 0.20  | 7.70 | 4.47  | 0.20    | 1       |

|     |      |      |      |      |   |
|-----|------|------|------|------|---|
| A20 | 0.81 | 7.72 | 3.61 | 0.81 | 1 |
| A21 | 1.02 | 7.85 | 5.00 | 1.02 | 1 |
| A22 | 0.20 | 7.70 | 4.47 | 0.20 | 1 |
| A23 | 1.29 | 7.87 | 4.24 | 1.29 | 1 |
| A24 | 0.20 | 7.70 | 4.47 | 0.20 | 1 |
| A25 | 6.95 | 1.00 | 8.31 | 1.00 | 2 |

At this stage, the cluster results are interpreted by providing information for each alternative. Cluster interpretation is based on the final centroid value. As seen, C1 → High Career Success, C2 → Medium Career Success, and C3 → Low Career Success. The interpretation can be seen in Table 7.

Table 7. Cluster Interpretation

| Alternative | Cluster | Information             |
|-------------|---------|-------------------------|
| A1          | 1       | High Career Success     |
| A2          | 1       | High Career Success     |
| A3          | 1       | High Career Success     |
| A4          | 1       | High Career Success     |
| A5          | 1       | High Career Success     |
| A6          | 1       | High Career Success     |
| A7          | 2       | Moderate Career Success |
| A8          | 1       | High Career Success     |
| A9          | 1       | High Career Success     |
| A10         | 1       | High Career Success     |
| A11         | 1       | High Career Success     |
| A12         | 1       | High Career Success     |
| A13         | 1       | High Career Success     |
| A14         | 1       | High Career Success     |
| A15         | 3       | Low Career Success      |
| A16         | 2       | Moderate Career Success |
| A17         | 1       | High Career Success     |
| A18         | 1       | High Career Success     |
| A19         | 1       | High Career Success     |
| A20         | 1       | High Career Success     |
| A21         | 1       | High Career Success     |
| A22         | 1       | High Career Success     |
| A23         | 1       | High Career Success     |
| A24         | 1       | High Career Success     |
| A25         | 2       | Moderate Career Success |

### 3.1.5. System Implementation

System implementation is the stage where the developed system is deployed and prepared for operation. At this stage, the interface design and program logic developed in the previous phase are implemented into program code using PHP version 8, the CodeIgniter 4 (CI4) Framework, and a MySQL/MariaDB database. The initial implementation process is conducted on a local server or localhost environment. The following section presents the implementation results of the proposed system. This page serves as the main access point for system security. Users are required to enter their registered username and password to authenticate and access the system according to their respective roles.

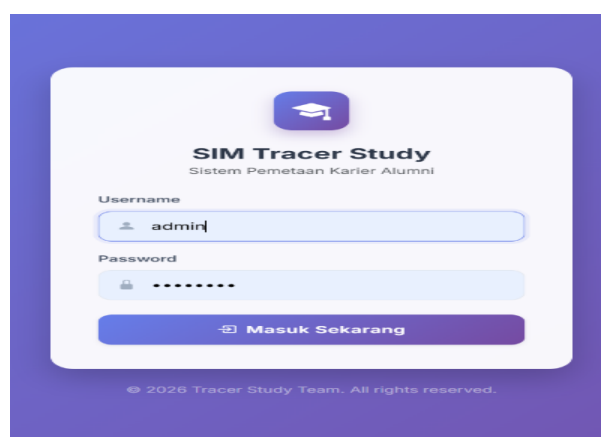


Figure 4. Login Page View

This page displays a summary of essential system information, including the total number of alternatives, attributes, and registered administrators. The dashboard provides an overview of the main data components managed within the system, allowing administrators to monitor the availability of data before conducting the clustering analysis process. This page also functions as an initial interface that helps users understand the current condition of the system and access available features efficiently.

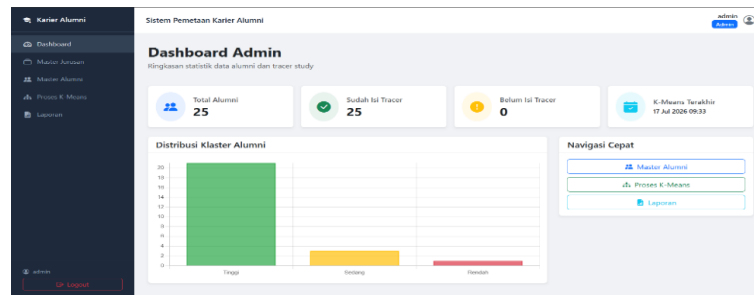


Figure 5. Home View

This page provides facilities for administrators to manage alumni data, including adding, editing, and deleting alumni records stored in the system. The page is designed to ensure that alumni information remains accurate, organized, and ready for further processing. Through this feature, administrators can update graduate data efficiently and maintain the completeness of information required for the tracer study and K-Means Clustering analysis process.

| # | NISN      | Nama           | Jurusan                | Tahun Lulus | Status Tracer | Aksi           |
|---|-----------|----------------|------------------------|-------------|---------------|----------------|
| 1 | 221011200 | Ahmad Solih    | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |
| 2 | 221011201 | Rayu Syahputra | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |
| 3 | 221011202 | Caca Handika   | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |
| 4 | 221011203 | Eka Ayu        | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |
| 5 | 221011204 | Frika Muliana  | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |
| 6 | 221011205 | Gulih Daryu    | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |
| 7 | 221011206 | Hanan Abanew   | Akuntansi dan Keuangan | 2022        | Sudah Isi     | [Edit] [Hapus] |

Figure 6. Alumni Master Data Management View

This page is used to input alumni questionnaire data based on the questions provided in the tracer study form. The collected data are stored in the system database and prepared for further analysis. This feature allows administrators to record alumni career information systematically, ensuring that the data are complete, accurate, and suitable for the K-Means Clustering process and graduate quality evaluation.

Figure 7. Tracer Study Data Input Display

This page displays the initial stage of the K-Means Clustering calculation process. The process begins by determining the number of clusters and selecting the initial centroid values as the basis for the clustering calculation. At this stage, the system prepares the required parameters and data attributes before performing distance calculations to identify the appropriate cluster membership for each data record.

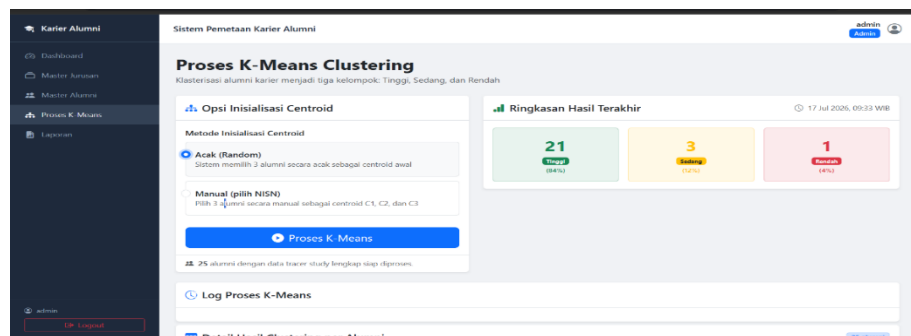


Figure 8. Initial Calculation View

This page displays the calculation process and the final clustering results generated by the K-Means method. The results include the grouping information and conclusions obtained from the clustering analysis. This feature allows users to review the iteration process, observe the distribution of alumni data across each cluster, and understand the characteristics of each group based on the selected career success indicators used in the analysis.

| #  | NISN      | Nama               | Jarak C1 | Jarak C2 | Jarak C3 | Kluster |
|----|-----------|--------------------|----------|----------|----------|---------|
| 1  | 221811308 | Ahmad Solih        | 8.1963   | 7.7608   | 6.8721   | Tinggi  |
| 2  | 221818773 | Andoko Simanjuntak | 5.1875   | 7.8740   | 4.2426   | Tinggi  |
| 3  | 230610008 | Arya Salih         | 6.1000   | 8.8603   | 6.4000   | Rendah  |
| 4  | 221818776 | Bagan Anzaka       | 8.1963   | 7.7608   | 6.8721   | Tinggi  |
| 5  | 221811301 | Bayu Syahputra     | 2.8886   | 7.7805   | 4.0000   | Tinggi  |
| 6  | 221811302 | Caca Handika       | 8.8189   | 7.7344   | 3.0000   | Tinggi  |
| 7  | 221811303 | Ika Ayu            | 1.4091   | 7.6852   | 6.1231   | Tinggi  |
| 8  | 221811304 | Fitika Mutiara     | 1.4091   | 7.6852   | 6.1231   | Tinggi  |
| 9  | 221811305 | Galth Dayu         | 8.9712   | 6.7862   | 6.1626   | Tinggi  |
| 10 | 221811306 | Hasan Abqary       | 4.1000   | 1.9461   | 5.0000   | Sedang  |
| 11 | 230610001 | Ihsanul Hakim      | 8.1963   | 7.7608   | 6.8721   | Tinggi  |
| 12 | 241800006 | Irfan Tri Putri    | 8.1963   | 7.7608   | 6.8721   | Tinggi  |
| 13 | 230610002 | Jaja Marwan        | 5.1875   | 7.8740   | 4.2426   | Tinggi  |

Figure 9. Calculation Results Display

This page is used to view, generate, and print reports containing the results of alumni career success clustering. The report provides detailed information about the cluster classification results and the characteristics of each alumni group. It can be used as supporting documentation for graduate quality evaluation and assist school management in reviewing alumni career patterns based on the results of the K-Means analysis.

| # | NISN      | Nama               | Jurusan                | Tahun Lulus | Status Aktivitas | Masa Tungggu | Th. Keberhasilan | Hasil Kluster |
|---|-----------|--------------------|------------------------|-------------|------------------|--------------|------------------|---------------|
| 1 | 221811308 | Ahmad Solih        | Akuntansi dan Keuangan | 2022        | Belum Aktif      | 0 hari       | Belum Berhasil   | Tinggi        |
| 2 | 221818773 | Andoko Simanjuntak | Biologi dan Pemasaran  | 2022        | Belum Aktif      | 0 hari       | Belum Berhasil   | Tinggi        |
| 3 | 230610008 | Arya Salih         | Akuntansi dan Keuangan | 2023        | Belum Aktif      | 0 hari       | Belum Berhasil   | Rendah        |
| 4 | 221818776 | Bagan Anzaka       | Biologi dan Pemasaran  | 2022        | Belum Aktif      | 0 hari       | Belum Berhasil   | Tinggi        |
| 5 | 221811301 | Bayu Syahputra     | Akuntansi dan Keuangan | 2022        | Belum Aktif      | 0 hari       | Belum Berhasil   | Tinggi        |
| 6 | 221811302 | Caca Handika       | Akuntansi dan Keuangan | 2023        | Belum Aktif      | 0 hari       | Belum Berhasil   | Tinggi        |

Figure 10. Printed View of Calculation Results Report

### 3.2. Discussion

The results indicate that K-Means can be used to map the career profiles of SMKN 1 Kisaran alumni in a more structured way. Converting categorical data into numerical form allowed the grouping process to run systematically. This aligns with the basic principle of data mining, which is to uncover patterns from large datasets so that scattered information can be read in a clearer structure (Han et al., 2022; Prasetyo, 2020). In this study, the selected attributes were employment status, waiting time before obtaining a job, and the match between job field and graduate competency. These attributes are relevant for describing the actual situation of alumni after graduation. At the initial stage, centroids were defined as reference points for distance calculation. Each alumni record was then measured using the Euclidean Distance formula to determine the nearest cluster. This process produced an initial grouping that was refined through further iterations. The movement of centroids across iterations shows that K-Means works adaptively until the clustering outcome becomes stable. Similar patterns have been reported in other studies using K-Means for student and social data grouping, since the algorithm separates records based on similarity without requiring prior labels (Fatkhudin et al., 2023; Novitasari et al., 2023).

Cluster 1 represents alumni with strong career success. Its main characteristics are a short waiting time for employment and a good match between the field of work and the competencies learned at school. This suggests that a portion of graduates is well prepared to enter the labor market. However, this result should be interpreted carefully. The success of one group does not automatically reflect the quality of all graduates. It is still necessary to examine whether this success is driven by the learning process, family support, industrial networks, or other external factors. If the school wants to maintain this positive pattern, curriculum alignment and industrial practice need to be sustained consistently (Ministry of Education and Culture of the Republic of Indonesia, 2021; Pratiwi & Mulyawan, 2023).

Cluster 2 reflects alumni with moderate career success. Some members of this group have already obtained jobs, but they experienced a longer waiting period or are working in fields that do not fully match their competencies. This finding suggests that the learning process has not fully bridged the needs of the labor market. It does not mean that the educational outcome is ineffective, but rather that there is still a gap that needs to be reduced. The school may need to review the proportion of practice, strengthen soft skills, and improve job readiness so that graduates can adapt more easily to workplace demands. In vocational education evaluation, data like this are useful as a basis for improvement rather than as a mere statistical report (Devie & Anamisa, 2020; Rudini, 2024).

Cluster 3 contains alumni who have not yet found employment. This group deserves closer attention because it reflects difficulties in the transition from school to the workplace. The condition may be related to limited experience, weak job readiness, or a mismatch between competencies and market needs. It may also be influenced by external conditions such as high competition in the labor market or limited access to job information. The school needs to respond seriously to this finding so that tracer study does not stop at data collection, but becomes the basis for follow-up strategies such as career guidance, mentoring, and expanded industrial partnerships (Ministry of Education and Culture of the Republic of Indonesia, 2021; Badan Pengembangan dan Pembinaan Bahasa, 2026).

From a methodological perspective, K-Means offers an advantage because the grouping results are obtained from clear calculations rather than subjective judgment. This makes the analysis easier to justify. However, clustering results still depend on the quality of the initial data. If alumni data are incomplete or respondents provide inconsistent answers, the grouping may deviate from the actual condition. Therefore, tracer study data quality must be maintained from the collection stage. Without well-structured data, even a sound method will not produce optimal results (Han et al., 2022; Wakhidah, 2020). The implementation of clustering results in a web-based system also shows the practical value of the study. The school does not only receive numbers and tables, but also a tool for reading alumni patterns more quickly. Dashboards, reports, and clustering outputs can help the school monitor graduate trends over time. From a management perspective, this supports decision-making based on evidence. However, such a system will only be effective if the school actually uses the results as a basis for evaluation, not merely as digital archives that are opened occasionally (Soufitri, 2023; Rudini, 2024).

#### 4. CONCLUSION

The application of the K-Means Clustering method in this study successfully grouped alumni career data from SMKN 1 Kisaran by transforming categorical data into numerical values that could be processed computationally. The clustering process was conducted by determining the initial centroid values, calculating the distance between each data point and the centroid using the Euclidean Distance method, and performing iterations until the centroid values reached a stable condition. The analysis results showed that the alumni data were classified into three main clusters. Cluster 1 represents alumni with a high level of career success, characterized by a shorter employment waiting period and a strong alignment between their occupational fields and competencies. Cluster 2 represents alumni with a moderate level of career success, indicated by a longer employment waiting period or partial alignment between occupational fields and competencies. Cluster 3 represents alumni with a lower level of career success, consisting mainly of alumni who have not yet obtained employment. Furthermore, the clustering results were implemented into a web-based information system that provides dashboard visualizations and analytical reports. The system output can be used by SMKN 1 Kisaran as supporting information for evaluating graduate quality and developing educational improvement strategies based on measurable alumni career data.

#### REFERENCES

- Arifah, S., Sweden, E. R., & Septian, M. R. D. (2023). Analisis Perbandingan Algoritma Clustering Dalam Melakukan Segmentasi Warna Pada Citra Jajan Tradisional. *Sebatik*, 27(1), 70-76. <https://doi.org/10.46984/sebatik.v27i1.2273>

- Badan Pengembangan dan Pembinaan Bahasa. (2026). Alumni. In Kamus Besar Bahasa Indonesia (KBBI). <https://kbbi.web.id/alumni>
- Devie, Y. K., & Anamisa, R. (2020). Sistem Informasi & Implementasi Untuk Pendukung Keputusan: Studi Kasus: Sitem Rekomendasi. Media Nusa Creative (MNC Publishing).
- Fatkhudin, A., Khambali, A., Artanto, F. A., & Zade, N. A. P. (2023). Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa. Jurnal Minfo Polgan, 12(1), 777-783. <https://doi.org/10.33395/jmp.v12i1.12494>
- FIKTI. (2023). K-Means clustering algorithm: Definition, function, and how it works. <https://fikti.umsu.ac.id/algoritma-k-meansclustering-pengertian-fungsi-dan-cara-kerja/>
- Han, J., Kamber, M., & Pei, J. (2022). Data mining: Concepts and techniques (4th ed.). Morgan Kaufmann.
- Harini, S., & Kusumawati, R. (2023). Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 di Indonesia Dengan Metode Davies-Bouldin Index (DBI). Jurnal Mnemonic, 6(2), 117-128. <https://doi.org/10.36040/mnemonic.v6i2.6642>
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2021). Tracer study sebagai instrumen evaluasi mutu pendidikan vokasi. Jakarta: Kementerian Pendidikan dan Kebudayaan Republik Indonesia.
- Novitasari, N., Nuris, N. D., & Herdiana, R. (2023). Penerapan Algoritma K-Means untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten di Jawabarat menggunakan Rapidminer. Jurnal Informatika Terpadu, 9(1), 68–73. <https://doi.org/10.54914/jit.v9i1.660>
- Prasetyo, E. (2020). Data mining: Concepts and applications using MATLAB. Andi Offset.
- Pratama, Y., Hendrawan, H., Rasywir, E., Carenina, B. T., & Anggraini, D. R. (2022). Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam. Building of Informatics, Technology and Science (BITS), 4(3), 1232–1240. <https://doi.org/10.47065/bits.v4i3.2549>
- Pratiwi, Y., & Mulyawan, M. (2023). Implementasi Algoritma K-Means untuk Menentukan Angka Harapan Hidup berdasarkan Tingkat Provinsi. Blend Sains Jurnal Teknik, 1(4), 284–294. <https://doi.org/10.56211/blendsains.v1i4.233>
- Quipper Campus. (2026). Alumni: Definition and meaning. <https://campus.quipper.com/kampuspedia/alumni>
- Ramadhan, R. F., Wijoyo, S. H., & Saputra, M. C. (2023). Penerapan Metode K-Means Clustering pada Ulasan Perumahan PT XYZ di Google Maps untuk Formulasi Strategi Bisnis dengan Analisis SWOT. (2023). Jurnal Pengembangan Teknologi Dan Ilmu Komputer, 7(6), 2879-2888. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12912>
- Rudini, A. (2024). Sistem informasi manajemen. CV Azka Pustaka.
- Soufitri, F. (2023). Konsep sistem informasi. PT Inovasi Pratama Internasional.
- Wakhidah, N. (2020). Clustering using the K-Means algorithm.