INVOTEK

Jurnal Inovasi Vokasional dan Teknologi

http://invotek.ppj.unp.ac.id/index.php/invotek

Vol. 25, No. 1, Feb 2025

ISSN: 1411 – 3411 (p)
ISSN: 2549 – 9815 (e)

Application of K-Means Data Mining to Measure the Level of Satisfaction of KITA Car Driving Course

Shelvina Ayu Wardani[1] *, Ruri Ashari Dalimunthe[2], Abdulkarim Syahputra[2]

[1] Information Systems Study Program, Royal University
Jl. Prof. H.M. Yamin No. 173, East Kisaran City District, Asahan Regency, North Sumatera, Indonesia-21222

[2] Computer Systems Study Program, Royal University
Jl. Prof. H.M. Yamin No. 173, East Kisaran City District, Asahan Regency, North Sumatera, Indonesia-21222

*Corresponding author: shelvinawardani14@gmail.com Doi: https://doi.org/10.24036/invotek.v25i1.1248

Abstract

KITA Driving Course, located in Kisaran, North Sumatra, faces challenges in maintaining service quality amidst increasing student enrollment. Despite operating since 2013, the institution lacks a structured, data-driven evaluation system to assess student satisfaction. This study applies a quantitative descriptive approach using the K-Means clustering algorithm to classify student satisfaction levels. Data were collected from 100 respondents through a questionnaire based on the five SERVQUAL dimensions: tangibles, reliability, responsiveness, assurance, and empathy. The K-Means algorithm grouped the satisfaction data into three categories: highly satisfied, satisfied, and dissatisfied. The majority of students were classified as satisfied (18 students), followed by highly satisfied (11 students), and dissatisfied (10 students). The findings indicate that the K-Means algorithm provides valuable insights into student satisfaction patterns, enabling targeted service improvements. The analysis also reveals that responsiveness and assurance were key areas of dissatisfaction, suggesting that instructor response time and safety assurance should be prioritized. These insights can help improve service strategies and can be adapted by other educational institutions for quality optimization through data analysis.

Keywords: Data Mining, K-Means Clustering, Satisfaction Level, Driving Course, Service Evaluation.

1. Introduction

The development of information technology has encouraged various sectors, including driving training services, to improve the quality of data-based services [1]. In the context of driving courses, the need for an objective evaluation system for student satisfaction is becoming increasingly important. KITA Driving Course, as one of the training institutions that continues to grow, faces challenges in understanding student expectations systematically. However, there is no analytical method that is able to group student satisfaction levels accurately and based on data. This creates a gap between the services provided and student needs, which can have an impact on the sustainability of services in the future.

Driving courses are a form of non-formal education that aims to improve a person's driving skills to be more skilled and confident [2]. In recent years, demand for driving course services has continued to increase along with the growth of private vehicle ownership in Indonesia [3], [4]. However, this growth has also given rise to a number of challenges that are common in various driving course institutions, including the absence of a structured participant satisfaction evaluation system, the gap between participant expectations and the quality of service received, and the limited use of data for decision making in improving services.

One of the institutions that provides driving course services is the KITA Driving Course, which is located on Jl. Cokroaminoto, Kisaran, and has been established since 2013. As one of the driving

Received date 11/03/2025, Revised date 14/04/2025, Accepted date 22/04/2025

P-ISSN: 1411-3414
E-ISSN: 2549-9815

courses that has been operating for a long time in the Kisaran area, the KITA Driving Course is the main choice for people who want to learn to drive a car safely and professionally. This institution offers various training packages designed for beginners to experienced drivers, with competent and experienced instructors.

As the number of course participants increases as shown in Table 1 below, KITA Driving Course faces challenges in maintaining the quality of teaching and service. Service quality is a major factor in building trust and satisfaction of course participants [5]. However, the increase in the number of participants can have an impact on their level of satisfaction if it is not balanced with continuous improvement in service quality.

Table 1. Number of Student Data 2024

No.	Month	Number of Students
1	January	32
2	February	27
3	March	30
4	April	25
5	May	21
6	June	29
7	July	32
8	August	26
9	September	30
10	October	25
11	November	39

Until now, KITA Driving Course has not had a structured system to measure the level of participant satisfaction objectively. Without documentation and analysis of participant feedback, service improvement efforts tend to be subjective and unfocused. Therefore, a quantitative approach is needed to analyze the level of participant satisfaction, so that it can be used as a basis for more effective service improvements (Table 2).

Table 2. Student Satisfaction Level Questionnaire

Category	Questionnaire
Tangibles (Physical Evidence)	Cleanliness of vehicles used in driving practice.
	Neatness and professionalism of the instructor's appearance.
	The quality of the vehicle used.
Reliability	The instructor's ability to provide training without errors or obstacles.
	Punctuality of course implementation according to the promised schedule.
Responsiveness	The instructor's speed in answering questions or giving directions.
	Staff's readiness to provide additional services or assistance as needed.
Assurance	Instructor's knowledge in teaching correct driving techniques.
	The level of safety felt by students during the process of learning to drive.
Empathy	The institution's understanding of students' learning needs in adjusting the training schedule.
	Friendly and caring attitude from the course institution.

To overcome these problems, this study applies a data mining method with the K-Means Clustering algorithm. Data mining is a technique that can be used to process data into useful knowledge, one of which is in analyzing the level of satisfaction of course participants [6]–[9]. The K-Means algorithm has the advantage of grouping data into several clusters based on certain similar characteristics [10]. Therefore, this method was chosen to group course participants into satisfaction categories such as

very satisfied, satisfied, and dissatisfied. With this method, KITA Driving Course is expected to be able to evaluate and improve the quality of its services more systematically and structured.

This study is based on the concept of data mining, especially the K-Means Clustering method, which has been proven effective in grouping data based on certain characteristics [10], [11]. Nugroho et al. [12] developed a driving course information system that showed a usability level of 85.8%, showing that information systems can improve the efficiency of course services. Ompusunggu et al. [8] applied the K-Means Clustering method in the analysis of driving course participant satisfaction, showing that this technique can help understand customer satisfaction patterns more objectively. Similarly, Kurniawan et al. [13] used K-Means to classify the level of education of the community, proving that this method can be used in various fields with structured data [8].

In addition, Rokhimakhumullah et al. [14] showed that data mining can be applied in the field of taxation, proving that machine learning- based analysis can support data-based decision making. Hermawan et al. [15] applied linear regression in the diagnosis of chronic kidney disease, showing that data analysis techniques can be used in various disciplines to improve prediction accuracy. This study adopts a similar approach by applying the K-Means Clustering method to group the satisfaction levels of driving course participants. Thus, this study aims to develop an information system that can automatically classify customer satisfaction levels based on the data collected, as well as assist course owners in data-based decision making.

Different from previous studies that have generally applied K-Means Clustering for broader educational or customer satisfaction analysis, this research makes a unique contribution by specifically focusing on the non-formal education sector, particularly driving course institutions—an area that has received limited attention in existing literature. This study applies K-Means Clustering to group satisfaction levels of students at the KITA Driving Course based on the five core dimensions of SERVQUAL (tangibles, reliability, responsiveness, assurance, and empathy), which are rarely used in combination with clustering in this specific context.

In addition, the research develops a data mining-based information system that can not only classify satisfaction levels automatically but also generate practical, data-driven recommendations for course managers. These recommendations are designed to bridge the gap between participant expectations and service delivery more effectively than traditional subjective evaluation methods. Thus, this study offers a novel, structured, and scalable approach to improving the quality of driving education services in Indonesia through intelligent data analysis.

2. Method

This research was conducted at the KITA Car Driving Course, located on Jl. Cokroaminoto, Kisaran, North Sumatra. The research implementation period was from November 2024 to March 2025. This study uses a descriptive quantitative approach, namely research that aims to measure and describe certain phenomena based on numerical data that can be calculated and analyzed statistically [16]. This study focuses on the level of satisfaction of course participants, which is analyzed using the SERVQUAL model.

2.1 Data Collection Techniques

Data collection was conducted using a closed questionnaire instrument distributed directly to course participants after they completed the last driving training session. The questionnaire was designed based on five main dimensions in the SERVQUAL model, namely: Tangibles (physical evidence), Reliability, Responsiveness (responsiveness), Assurance (guarantee), and Empathy (empathy)

The data obtained from the questionnaire will be processed using the K-Means Clustering method to group the level of satisfaction of course participants into a Likert scale with a range of 1–5 (Table 3), where [17]:

Table 3. Assessment Indicators

No.	Rating Scale	Weight
1	Very Satisfied (SP)	5
2	Satisfied (P)	4
3	Quite Satisfied (CP)	3

4	Not Satisfied (KP)	2
5	Not Satisfied (TP)	1

The rating scale used in the questionnaire uses a range of 1 to 5, where 1 indicates dissatisfaction and 5 indicates very high satisfaction. This scale is then coded into a numeric number (1, 2, 3, 4, 5) for each questionnaire question, which will be used in the clustering analysis. The results of this clustering will provide an overview of the level of participant satisfaction and the factors that influence it.

Each of the five SERVQUAL dimensions was operationalized into specific indicators, measured using a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The indicators are as follows (Table 4):

Table 4. Student Satisfaction Level Questionnaire

Category	Questionnaire
Tangibles (Physical Evidence)	Cleanliness of vehicles used in driving practice.
	Neatness and professionalism of the instructor's appearance.
	The quality of the vehicle used.
Reliability	The instructor's ability to provide training without errors or obstacles.
	Punctuality of course implementation according to the promised schedule.
Responsiveness	The instructor's speed in answering questions or giving directions.
	Staff's readiness to provide additional services or assistance as needed.
Assurance	Instructor's knowledge in teaching correct driving techniques.
	The level of safety felt by students during the process of learning to drive.
Empathy	The institution's understanding of students' learning needs in adjusting the training schedule.
	Friendly and caring attitude from the course institution.

The indicators used in the questionnaire are detailed in Table 3 and categorized under the respective SERVQUAL dimensions as follows:

- Tangibles: Cleanliness of vehicles used in driving practice; neatness and professionalism of the instructor's appearance; and the quality of vehicles used.
- Reliability: Instructor’s ability to deliver lessons without error; punctuality of the schedule.
- Responsiveness: Instructor’s speed in answering questions or giving directions; staff’s readiness to offer additional help when needed.
- Assurance: Instructor’s knowledge in teaching proper driving techniques; students’ perceived safety during training.
- Empathy: Flexibility in adapting to students’ schedules; friendly and attentive attitudes from staff.

The collected data was then processed using the K-Means Clustering algorithm to group participants based on their overall satisfaction levels.

2.2 Sampling Determination Techniques

Determination of the number of samples is done by purposive sampling method, namely sampling based on certain criteria. In this case, the criteria for participants selected are: a) Have attended at least 5 training sessions, b) Have completed the entire course program, c) Willing to fill out the questionnaire.

The purposive sampling method was chosen because it aims to ensure that the participants who are sampled have sufficient experience in taking the course, so that they can provide more accurate and valid feedback regarding their satisfaction with the course service [18], [19]. The criteria for participants who have attended at least 5 training sessions are expected to provide a more comprehensive picture of the quality of the course service, as well as increase the reliability of the research results.

Based on the 2024 course data (see Table 1), the total number of participants until November was 316 people. From this number, a sample of 39 participants was determined to represent the population of quantitative descriptive analysis (Table 5).

P-ISSN: 1411-3414
E-ISSN: 2549-9815

Table 5. Questionnaire Data

#	Questionnaire Statement
K1	Cleanliness of vehicles used in driving practice.
K2	Neatness and professionalism of the instructor's appearance.
K3	The quality of the vehicle used.
K4	The instructor's ability to provide training without errors or obstacles.
K5	Punctuality of course implementation according to the promised schedule.
K6	The instructor's speed in answering questions or giving directions.
K7	Staff's readiness to provide additional services or assistance as needed.
K8	Instructor's knowledge in teaching correct driving techniques.
K9	The level of safety felt by students during the process of learning to drive.
K10	The institution's understanding of students' learning needs in adjusting the training schedule.
K11	Friendly and caring attitude from the course institution.

Table 6 shows the results of the questionnaire data normalization, where the values given by participants (e.g. 'Very Satisfied' or 'Satisfied') are converted into numbers to facilitate data processing using the K-Means algorithm.

Table 6. Questionnaire Result Data

#	Alternative	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11
1	A1	P	SP	P	CP	P	CP	P	SP	P	CP	P
2	A2	SP	P	P	P	P	SP	P	P	P	SP	SP
3	A3	CP	P	P	SP	P	P	CP	CP	CP	CP	P
4	A4	SP	SP	SP	SP	CP	SP	SP	P	P	SP	P
5	A5	SP	SP	SP	SP	SP	SP	SP	SP	SP	SP	SP
…	…	…	…	…	…	…	…	…	…	…	…	…
34	A34	P	P	P	P	P	P	SP	P	P	P	SP
35	A35	P	P	P	SP	P	P	P	P	P	P	SP
36	A36	SP	SP	SP	SP	SP	SP	SP	SP	P	P	SP
37	A37	CP	P	P	P	CP	P	CP	P	CP	P	SP
38	A38	SP	SP	P	CP	P	P	P	CP	P	P	SP
39	A39	P	SP	P	SP	P	SP	SP	P	P	P	SP

Table 7. Normalization of Questionnaire Result Data

#	Alternative	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11
1	A1	4	5	4	3	4	3	4	5	4	3	4
2	A2	5	4	4	4	4	5	4	4	4	5	5
3	A3	3	4	4	5	4	4	3	3	3	3	4
4	A4	5	5	5	5	3	5	5	4	4	5	4
5	A5	5	5	5	5	5	5	5	5	5	5	5
…	…	…	…	…	…	…	…	…	…	…	…	…
34	A34	4	4	4	4	4	4	5	4	4	4	5
35	A35	4	4	4	5	4	4	4	4	4	4	5
36	A36	5	5	5	5	5	5	5	5	4	4	5
37	A37	3	4	4	4	3	4	3	4	3	4	5
38	A38	5	5	4	3	4	4	4	3	4	4	5
39	A39	4	5	4	5	4	5	5	4	4	4	5

Once the data has been normalized (Table 7), the K-Means Clustering algorithm is applied to group course participants based on their satisfaction levels. The algorithm categorizes participants into three main clusters: Very Satisfied, Satisfied, and Less Satisfied. Each cluster is then analyzed to identify the service factors contributing to the respective satisfaction level. This analysis enables course

managers to pinpoint areas that require improvement or reinforcement, ultimately helping to enhance the overall quality of service.

2.3 Research Framework

The following is a description of the research framework that has been compiled systematically and clearly (Figure 1):

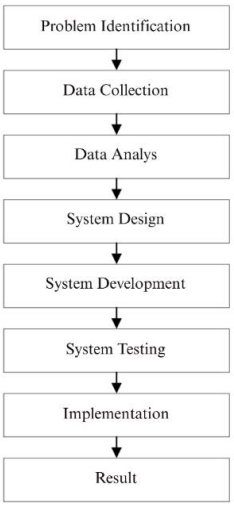


Figure 1. Research Framework

This study began with problem identification, focusing on the absence of a structured system for measuring student satisfaction at the KITA Driving Course. Data was collected through interviews and observations, then analyzed to identify existing obstacles and propose suitable solutions. The system design phase included interface design, database structuring, and programming. Development was carried out using software such as Visual Studio 2010, Crystal Reports, and XAMPP. System testing was performed using the black-box method to ensure that all functionalities worked as intended. After implementation, the system was applied in a real-world setting, and the results demonstrated its effectiveness in objectively measuring student satisfaction.

3. Results and Discussion

The next step is to evaluate the steps involved in grouping data into clusters based on similar characteristics. The k-means algorithm works by grouping data into k clusters, where *k* is the number of predetermined clusters.

3.1 Data Analysis

The process begins with a random centroid initialization for each cluster. Each data in the dataset is then measured for its distance to the nearest centroid and assigned to that cluster (Table 8). The manual calculation process of the k-means algorithm in the application of k-means data mining to measure the level of satisfaction of the KITA Car Driving Course is as follows:

- Determining the Number of Clusters
  The number of clusters has been determined to be 3 clusters, namely C1 Very Satisfied, C2 Satisfied and C3 Not Satisfied.
- Determining the Initial Center of the Centroid
  To determine the initial center (centroid) is determined from existing data. For the initial determination it is assumed: a) The 5th data is taken, namely A5 as the center of Cluster 1, b) The 15th data is taken, namely A15 as the center of Cluster 2. C) The 37th data is taken, namely A37 as the center of Cluster 3.

Table 8. Initial Centroid Center

Item	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11
1st Cluster	SP	SP	SP	SP	SP	SP	SP	SP	SP	SP	SP
2nd Cluster	P	P	P	SP	SP	P	P	P	P	SP	P

3rd Cluster	CP	P	P	P	CP	P	CP	P	CP	P	SP

The selection of the 5th, 15th, and 37th data points as the initial centroid centers aims to initiate the clustering process with data that exhibit diverse characteristics. These points are selected randomly, while still considering the variation within the dataset to ensure better representation in each cluster. This approach helps the algorithm group participants more effectively. Following the selection of these initial centroids, normalization is performed, resulting in the following values (Table 9):

Table 9. Normalization of Cluster Values

Item	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11
1st Cluster	5	5	5	5	5	5	5	5	5	5	5
2nd Cluster	4	4	4	5	5	4	4	4	4	5	4
3rd Cluster	3	4	4	4	3	4	3	4	3	4	5

Normalization of values is carried out to standardize the assessment scale provided by participants, ensuring that variations in each satisfaction category can be analyzed fairly. This normalization process enhances the consistency of the data used in calculating the distances between points within each cluster. The calculation was carried out in 3 (three) iterations, with the calculation results as follows:

A1 = ('3.42' , '2.78' , '2.32')
A1 = ('C1' , 'C2' , 'C3')
The result of the shortest distance cluster value is 2.32 in cluster 3.
A2 = ('1.85' , '1.47' , '2.89')
A2 = ('C1' , 'C2' , 'C3')
The result of the shortest distance cluster value is 1.47 in cluster 2.
A3 = ('4.19' , '2.80' , '1.99')
A3 = ('C1 , 'C2' , 'C3')
The result of the shortest distance cluster value is 1.99 in cluster 3.

Then until the cluster results at A39, the shortest distance results for the 3rd iteration are obtained in the Table 10 below.

Table 10. Shortest Distance of 3rd Iteration

#	Alternative	Shortest Distance	Square	Cluster
1.	A1	2.32	5.36	3
2.	A2	1.47	2.15	2
3.	A3	1.99	3.96	3
4.	A4	2.08	4.32	1
5.	A5	1.23	1.50	1
6.	A6	1.50	2.26	2
7.	A7	2.59	6.70	2
8.	A8	2.23	4.96	3
9.	A9	1.35	1.81	2
10.	A10	1.69	2.87	1
11.	A11	1,12	1.26	2
12.	A12	1.58	2.48	2
13.	A13	1.84	3.37	2
14.	A14	1.71	2.93	2
15.	A15	1.61	2.59	2
16.	A16	1.46	2.14	1
17.	A17	2.36	5.56	3
18.	A18	1.64	2.70	2
19.	A19	1.83	3.36	3
20.	A20	2.34	5.48	2

#	Alternative	Shortest Distance	Square	Cluster
21.	A21	1.77	3.15	2
22.	A22	1.87	3.50	1
23.	A23	1.12	1.26	2
24.	A24	1.30	1.69	1
25.	A25	1.66	2.76	3
26.	A26	1.89	3.56	3
27.	A27	1.58	2.50	1
28.	A28	1.11	1.23	1
29.	A29	1.87	3.48	2
30.	A30	1.30	1.69	1
31.	A31	1.58	2.50	1
32.	A32	1.99	3.96	3
33.	A33	1.60	2.56	3
34.	A34	1.22	1.48	2
35.	A35	0.90	0.81	2
36.	A36	1.23	1.50	1
37.	A37	1.54	2.36	3
38.	A38	2.34	5.48	2
39.	A39	1.39	1.93	2
	WCV		57.76	

The ratio of the magnitude between BCV (Between Cluster Variation) and WCV (Within Cluster Variation). ecause:

Centroid m1 = (4.818), (4.636), (4.636), (4.636), (4.455), (4.636), (5.000), (4.818), (4.273), (4.727), (4.909)

Centroid m2 = (4.222), (4.278), (3.833), (4.556), (4.056), (4.444), (4.222), (4.222), (3.944), (4.222), (4.667)

Centroid m3 = (3.400), (3.900), (3.700), (3.800), (3.400), (3.600), (3.800), (3.800), (3.300), (3.600), (4.100)

BCV = d(m1,m2) + d(m1,m2) + d(m1,m2) = 7.069

Ratio = BCV/WCV

Ratio = 7.069/115.19 = 0.0614

Since the 3rd and 4th iterations show that the group members and the ratio size are stable and there is no significant change in the cluster center (centroid), the iteration process is stopped. This shows that the clustering results have achieved stability and convergence. There are three categories of student satisfaction grouped into three clusters:

- Very Satisfied (C1) – 11 students:
  The participants were: Joshua Estefan Silitonga, Nur Rizky Aisyah Silalahi, Eka Puspita, Wira Sukmana, Azura Alni Novalni, Novia Zesika Nainggolan, M. Rifky Arfizan Guming, Darmawan Saragih, Johny, Haykal Raynaldi, and Yowan Putra.
- Satisfied (C2) – 18 students:
  The following are the names of the participants:
- Not Satisfied (C3) – 10 students:
  The following are the names of the participants: Rizal Azis, Kim Guek, Ruh, Zahara Oktavia, Nur Aisyah Siregar, Agustian, Ujang Gunawan, Suliana, Apri, and Rudy Chandra Siregar.

The clustering results show that the majority of students are in the Satisfied category (C2) with 18 students, while the Very Satisfied category (C1) has 11 students and Dissatisfied (C3) has 10 students. The iteration process was stopped after the 4th iteration, because in the 3rd and 4th iterations, the cluster members and ratios were stable, indicating that the K-Means algorithm had reached convergence. This algorithm will continue to cluster the data until the difference between iterations is getting smaller or there is no significant change in the cluster center (centroid) [20], [21].

However, to provide a deeper understanding of the factors that influence satisfaction, it is important to conduct additional analysis of the dimensions that influence each satisfaction category.

P-ISSN: 1411-3414
E-ISSN: 2549-9815

Based on the SERVQUAL model, the Dissatisfied group (C3) shows a possible deficiency in the Responsiveness and Assurance dimensions, indicating that students expect a faster instructor response and better safety assurance during training. This is in line with the research of Ompusunggu et al. [8], which found that in driving courses, the instructor's response speed and vehicle conditions greatly affect the level of participant satisfaction.

For the Satisfied (C2) and Very Satisfied (C1) groups, other factors such as instructor quality, vehicle comfort, and supporting facilities may play a more dominant role. Therefore, this study suggests conducting further analysis of other variables, such as the quality of teaching materials and instructor interaction, which can clarify the reasons for participant satisfaction. Based on the results of the study, here are some recommendations that can be implemented by KITA Driving Course

When compared to the research of Nugroho et al. [12], which obtained a satisfaction level of 85.8% on the technology-based driving course information system, the results of this study indicate that the use of an application-based feedback system may be able to further increase student satisfaction. Therefore, as a follow-up, driving courses can consider using a technology-based evaluation system to increase participant satisfaction.

3.2 System Implementation

The proposed new information system aims to examine and evaluate the flow of information within the company after its implementation. The proposed system in this study is as follows (Figure 2):

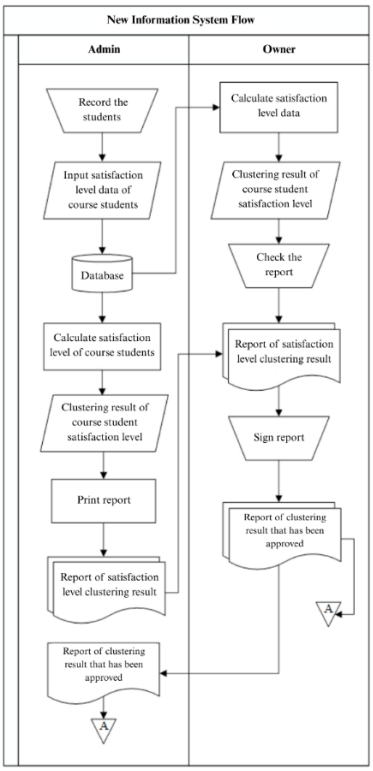


Figure 2. Research Framework

In an effort to improve the quality of driving course services, an evaluation of the level of student satisfaction is needed. This evaluation is carried out by collecting data from questionnaires filled out by students, then analyzed to determine their level of satisfaction. With a structured information system, this process can be carried out more effectively and accurately. The proposed system allows the admin to record student data and questionnaire results, input data into the system, and calculate the level of satisfaction. After the clustering results appear, the report will be printed and given to the driving course owner for further analysis.

The implementation process includes collecting student satisfaction data, processing data using the K-Means Clustering technique, and testing the system that has been developed using Microsoft Visual Studio 2010 and MySQL as a database. The results of this implementation can help the KITA Driving Course in identifying the level of student satisfaction more objectively and systematically, so

P-ISSN: 1411-3414
E-ISSN: 2549-9815

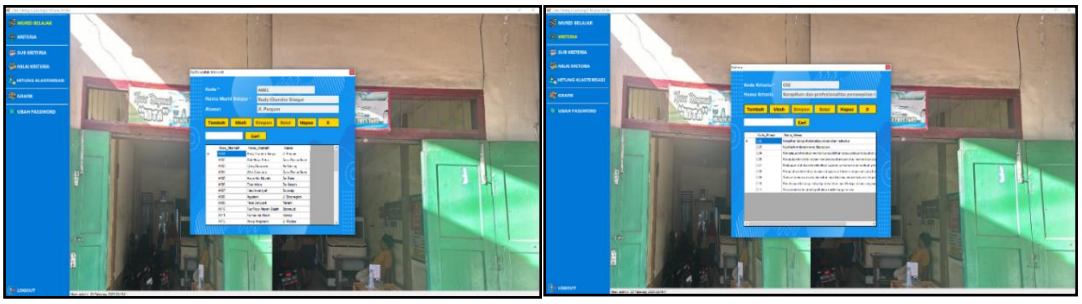
that it can be used as a basis for improving the quality of course services. The system interface includes several main features, such as student data input pages, satisfaction questionnaire forms, K-Means clustering results displays, and visualization graphs of analysis results. With the implementation of an intuitive and user-friendly interface, the course can easily monitor and evaluate the level of student satisfaction in real-time, thus supporting decision making in improving the quality of course services.

The following are some views of the system design results that have been explained in the previous section. The login page is the first page accessed by the admin. The admin must have valid credentials before they can access all features in the application (Figure 3a). The main menu page consists of the main features found in the application. This page can be accessed when the admin has valid credentials (Figure 3b).

a. b.

Figure 3. a) Login Page, b). Main Menu Page

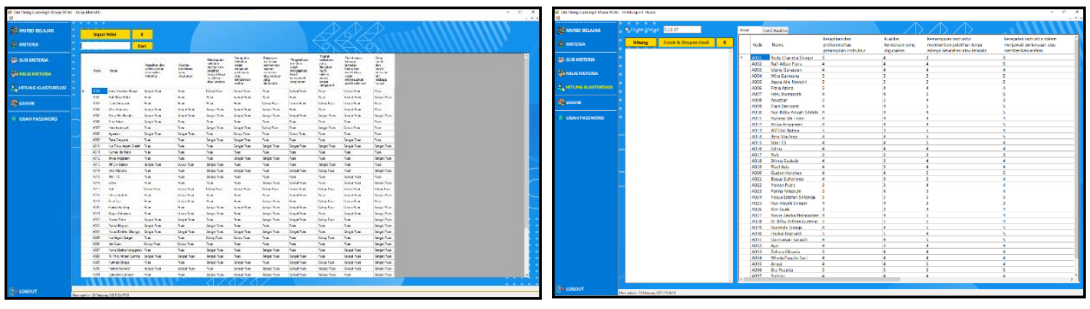
The student learning data page contains features for managing course participant data. Admin can add data, edit or delete student data (Figure 4a). The criteria page consists of features in managing criteria data. On this page, the admin determines what criteria will be used in the calculation. In this research case, the criteria page contains the SERVQUAL criteria (Figure 4b).



a. b.

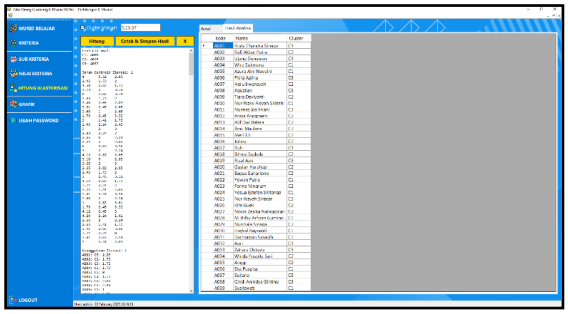
Figure 4. a) Student Data Page, b) Criteria Page

The criteria value page contains the values that have been set from previously determined criteria. On this page, the admin can add values, change, or delete criteria values (Figure 5a). The criteria calculation page consists of functions related to the calculation in the K-Means method discussed in the study. This page has a calculate button that functions to start the calculation based on previously set data and criteria (Figure 5b). The results of the calculations on the calculation page will be displayed on the calculation results page. This page contains all the results, iterations, values, and alternative data used in the calculation (Figure 5c).



a. b.

P-ISSN: 1411-3414
E-ISSN: 2549-9815



c

Figure 5. a) Criteria Value Page, b) Clustering Calculation Page, c) Calculation Results Page

4. Conclusion

Based on the research findings, it can be concluded that the application of data mining clustering, particularly the K-Means method, in measuring student satisfaction at the KITA Car Driving Course has proven effective in providing a more systematic and objective analysis. The K-Means method successfully grouped student satisfaction levels into three main clusters: Cluster C1 (Very Satisfied), Cluster C2 (Satisfied), and Cluster C3 (Not Satisfied), offering a deeper understanding of student perceptions of the course services. Moreover, the results demonstrate consistency with manual data analysis, confirming that the K-Means clustering method not only enhances accuracy in the clustering process but also reduces the potential for subjective errors inherent in manual analysis. However, it is important to acknowledge that potential biases in data clustering—such as imbalances in the number of data points per cluster and the selection of initial centroid positions—can influence the results. These biases were addressed through multiple iterations and testing to ensure the outcomes are both representative and reliable. This study illustrates that data mining techniques such as K-Means clustering can be effectively applied in the education and training sector globally, particularly for analyzing customer or participant satisfaction. Such approaches can offer valuable insights for training service providers worldwide to enhance service quality and optimize user experience. Therefore, this system can serve as an effective tool for the KITA Driving Course to evaluate and improve its service quality, as well as to maintain and increase student satisfaction over time, with potential applicability across various industry sectors.

References

[1] P. S. Aithal and S. Aithal, "Management of ICCT underlying Technologies used for Digital Service Innovation," *International Journal of Management, Technology, and Social Sciences (IJMTS)*, no. 4, pp. 110–136, 2019, [Online]. Available: https://ssrn.com/abstract=3512944

[2] V. Beanland, N. Goode, P. M. Salmon, and M. G. Lenné, "Is there a case for driver training? A review of the efficacy of pre- and post-licence driver training," *Saf Sci*, vol. 51, no. 1, pp. 127–137, Jan. 2013, doi: 10.1016/J.SSCI.2012.06.021.

[3] N. Made, D. Elianti, D. Putra Githa, A. A. Ngurah, and H. Susila, "Android-Based Driving Course Information System," *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, vol. 10, no. 1, pp. 24–33, Nov. 2021, doi: 10.24843/JIM.2022.V10.I01.P03.

[4] R. Ruslan, S. Sugiarto, R. Anggraini, and S. M. Saleh, "Forecasting Private Vehicle Ownership and It Effect to Public Transportation Planning in Banda Aceh, Indonesia," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Sep. 2020. doi: 10.1088/1757-899X/917/1/012040.

[5] J. S. Chiou and C. Droge, "Service quality, trust, specific asset investment, and expertise: Direct and indirect effects in a satisfaction-loyalty framework," *J Acad Mark Sci*, vol. 34, no. 4, pp. 613–627, Oct. 2006, doi: 10.1177/0092070306286934/METRICS.

[6] M. Nilashi *et al.*, "What Factors Influence Students Satisfaction in Massive Open Online Courses? Findings from User-Generated Content Using Educational Data Mining," *Educ Inf Technol (Dordr)*, vol. 27, no. 7, pp. 9401–9435, Aug. 2022, doi: 10.1007/S10639-022-10997-

7/TABLES/8.

[7] K. Dejaeger, F. Goethals, A. Giangreco, L. Mola, and B. Baesens, "Gaining insight into student satisfaction using comprehensible data mining techniques," *Eur J Oper Res*, vol. 218, no. 2, pp. 548–562, Apr. 2012, doi: 10.1016/J.EJOR.2011.11.022.

[8] N. A. Ompusunggu, P. Purwadi, and A. Calam, "Penerapan Data Mining Untuk Mengukur Tingkat Kepuasan Pada Peserta Kursus Mengemudi Mobil" Karunia" Medan Menggunakan Metode K-Means Clustering," *J. Cyber Tech*, vol. 3, no. 8, pp. 1–9, 2020.

[9] B. Oralhan, K. Uyar, and Z. Oralhan, "Customer Satisfaction Using Data Mining Approach," *International Journal of Intelligent Systems and Applications in Engineering*, no. 4, pp. 63–66, 2016, [Online]. Available: www.atscience.org/IJISAE

[10] R. C. de Amorim, "A survey on feature weighting based K-Means algorithms," *Journal of Classification (Springer)*, Sep. 2015, [Online]. Available: http://arxiv.org/abs/1601.03483

[11] R. Liu, "Data Analysis of Educational Evaluation Using K-Means Clustering Method," *Comput Intell Neurosci*, vol. 2022, pp. 1–10, 2022, doi: 10.1155/2022/3762431.

[12] N. Nugroho, Y. Rahmanto, D. Alita, J. Z. Pagar Alam, and L. Ratu, "Software Development Sistem Informasi Kursus Mengemudi (Kasus: Kursus Mengemudi Widi Mandiri)," 2021.

[13] R. Kurniawan and M. Mukarrobin, "Jurnal Ilmiah 'Technologia' KLASTERISASI TINGKAT PENDIDIKAN DI DKI JAKARTA PADA TINGKAT KECAMATAN MENGGUNAKAN ALGORITMA K-MEANS." [Online]. Available: https://data.jakarta.go.id.

[14] Dewi Noor Fatikhah Rokhimakhumullah, Devi Nur Cahaya Ningsih, Aji Widya Firmansyah, and Ayun Hanafiyah, "ANALISIS POTENSI E-COMMERCE MELALUI IMPLEMENTASI DATA MINING DALAM PERPAJAKAN: SEBUAH STUDI KOMPARASI," *Juremi: Jurnal Riset Ekonomi*.

[15] A. K. Hermawan, A. Nugroho, and Edora, "Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier," *Bull. Inf. Technol.*, vol. 4, no. 1 SE-Articles, Mar. 2023, doi: 10.47065/bit.v4i1.475.

[16] H. K. Mohajan, "Munich Personal RePEc Archive Quantitative Research: A Successful Investigation in Natural and Social Sciences," *Journal of Economic Development, Environment and People*, vol. 4, no. 9, 2020.

[17] Topal, A. Deveci, and A. K. Geçer, "Examination of Student Satisfaction with e-courses by Clustering Analysis," *Innoeduca. International Journal of Technology and Educational Innovation*, vol. 9, no. 2, pp. 39–50, 2023.

[18] F. Nyimbili and L. Nyimbili, "Types of Purposive Sampling Techniques with Their Examples and Application in Qualitative Research Studies," *British Journal of Multidisciplinary and Advanced Studies*, vol. 5, no. 1, pp. 90–99, Feb. 2024, doi: 10.37745/bjmas.2022.0419.

[19] J. Omona, "Sampling in Qualitative Research: Improving the Quality of Research Outcomes in Higher Education," *Makerere Journal of Higher Education*, vol. 4, no. 2, Jul. 2013, doi: 10.4314/majohe.v4i2.4.

[20] S. Zahra, M. A. Ghazanfar, A. Khalid, M. A. Azam, U. Naeem, and A. Prugel-Bennett, "Novel centroid selection approaches for KMeans-clustering based recommender systems," *Inf Sci (N Y)*, vol. 320, pp. 156–189, Nov. 2015, doi: 10.1016/j.ins.2015.03.062.

[21] E. W. Liyanto, A. Homaidi, and A. Lutfi, "IMPLEMENTASI K-MEANS CLUSTERING MENGGUNAKAN RAPIDMINER DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI PROVINSI JAWA TIMUR," vol. 19, pp. 205–216, 2024.