X-Means Clustering for UX Evaluation of the Candy CBT Application Using the SUS Instrument

Laila Nurzanah[1], Ilka Zufria[2], Suwannit Chareen Chit[3]

[1,2]*Computer Science Department, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia*

[3]*School of Computing Department, Northern University of Malaysia, Malaysia*

[1]lailanurzanahh@gmail.com (*)

[2]ilkazufria@uinsu.ac.id

[3]chareen@gmail.com

Received: 2025-12-04; Accepted: 2026-01-17; Published: 2026-01-29

Abstract— The Candy CBT application at *SMKS Pancabudi Medan* has been operating for three years without adequate usability reviews, posing a high risk of serious problems in user experience (UX), especially in terms of system responsiveness. Therefore, this study aims to analyze the level of system usability and group users based on three aspects (GUI, Navigation, and Responsiveness) using X-Means Clustering to develop recommendations for improvement. The method used is a data mining-based evaluative study involving 400 respondents of a modified System Usability Scale (SUS) questionnaire, with data processed through reverse scoring and Z-Score Normalization before clustering. The results show that the GUI (4.53) and Navigation (4.51) aspects are rated very good. Still, responsiveness is very low (1.48), becoming a major weakness consistent across all clusters. The X-Means Clustering model automatically determined two optimal clusters, with the most dominant cluster specifically showing extreme dissatisfaction with responsiveness (score 1.08). This study contributes a more granular usability evaluation approach by integrating aspect-based SUS features with X-Means Clustering, enabling more precise identification of critical UX weaknesses and supporting data-driven prioritization of usability improvements in CBT systems. Therefore, the application's usability needs to be significantly improved, and recommendations should focus on enhancing non-functional requirements, particularly system performance and stability, to ensure better multi-device responsiveness, with the cluster showing the highest dissatisfaction as the top priority.

Keywords— X-Means Clustering; User Experience; UX; System Usability Scale; SUS; CBT Application; Responsiveness; Data Mining.

I. Introduction

The adoption of Computer-Based Testing (CBT) applications has become a recognized strategy for improving the efficiency and accuracy of educational evaluation [1][2]. The ongoing implementation of this technology demands rigorous usability evaluation. The CBT Candy application at SMKS Pancabudi Medan, the object of study, has been in operation for 3 years without adequate review. This condition is at high risk of causing serious usability issues, such as navigation confusion, poor graphical interface design, and poor user responsiveness [3]. These issues directly affect the user experience (UX) and the credibility of the exam system. Therefore, the evaluation was carried out using data mining to analyze data from the System Usability Scale (SUS) questionnaire completed by application users [4][5].

To analyze system usability problems, the System Usability Scale (SUS) is used as a standard instrument [6][7]. The effectiveness of this measurement was enhanced by applying clustering techniques to group user responses, with X-Means Clustering proven to be an effective method for stability analysis [8]. As a relevant foundation research, [9] evaluated the usability of the STMIK IKMI Cirebon Virtual Tour website by applying X-Means Clustering to SUS questionnaire results. The study successfully identified user groups with distinct characteristics, revealing that the optimal clustering consisted of two user clusters, with Cluster 0 achieving the best performance, indicated by an average cluster distance of 1.093 and a Davies–Bouldin Index (DBI) value of 0.296.

While the study in [9] demonstrated the effectiveness of X-Meanfor segmenting users, the clustering process relied solely on the overall SUS score, resulting in user groups that primarily reflected general usability perceptions. This condition indicates a methodological limitation, not in the number of clusters obtained, but in the level of analytical granularity achieved byhe selected features [10]. Relying solely on aggregate usability scores limits the exploration of deeper insights into diverse user experience profiles. Moreover, previous studies tended to focus on general usability outcomes and did not sufficiently examine critical user experience (UX) dimensions, such as Graphical User Interface (GUI), navigation, and responsiveness. The absence of a detailed analysis of these aspects makes it difficult to identify the key drivers of user satisfaction and dissatisfaction. Therefore, a more decomposed feature representation is required to provide a clearer and more interpretable diagnosis of system usability and user experience quality [10].

To overcome these limitations, this study distinguishes clustering analysis by applying the X-Means algorithm to produce two distinct clusters. This group is based on the average score of the GUI, Navigation, and Responsiveness aspects. This cluster differentiation approach is a methodological novelty that aims to provide a comprehensive understanding of the user experience and identify the most critical and impactful areas of improvement [11]. These three

DOI : https://doi.org/10.25139/inform.v11i1.11277

aspects were deliberately selected because they represent the most critical functional and non-functional usability dimensions in Computer-Based Test (CBT) systems, as consistently emphasized in prior usability and UX evaluation studies. In addition, this study will consider integrating other methods into the clustering process to improve the accuracy and quality of analysis results [12][13].

The purpose of this study is to analyze the usability of the CBT Candy application using the SUS questionnaire, apply the X-Means Clustering algorithm to group users based on GUI, navigation, and responsiveness, and compile improvement recommendations to enhance quality and user satisfaction. This research is expected to provide benefits by developing a more granular SUS data clustering analysis model and by serving as a guide for developers to specify more comprehensive non-functional requirements thatmprove system stability, ensure ease of use, and support real-time exam management by teachers [14].

II. Research Methodology

A. Research Design

This study uses an evaluative study based on the data mining method. The research design focuses on assessing the quality of User Experience (UX) of the CBT Candy application through user perception data analysis [15][16].

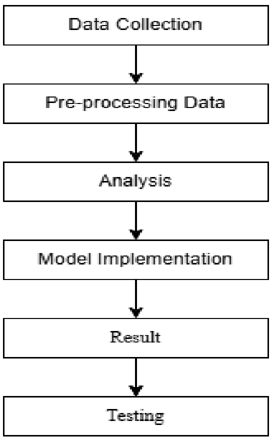

Fig.1. Research Stages

The research workflow is structured into six sequential stages in Fig.1: Data Collection, Data Pre-processing, Analysis, Model Implementation, Results, and Testing. This design is intended to ensure rigorous data analysis and generate data-driven improvement recommendations.

B. Subject and Location

The CBT Candy application is the object of the research, and *SMKS Pancabudi Medan* is the designated research location, as it provides a context for the use of relevant and sustainable applications for approximately 3 years [17][18].

The study included active users of the CBT Candy app, with a sample of 400 respondents. The research population consists of all relevant application users. The sampling technique used is Purposive Sampling, ensuring respondents have hands-on, adequate experience with the evaluated application.

C. Research Instrument

The System Usability Scale (SUS) questionnaire is the main instrument. This questionnaire is modified and tested for validity (Instrument Test) before deployment. The instrument consists of 10 statements, and respondents rate each statement on a Likert scale from 1 to 5. The variable measured is the level of usability, which is described in three aspects as feature inputs for modelling: Graphical User Interface (GUI), Navigation, and Responsiveness [19][20]. The Validity Test of the instrument is performed to ensure that all statement items meet the valid correlation requirements.

D. Data Collection

Data collection was carried out by distributing validated SUS questionnaires to 400 respondents. All collected Likert-scale responses form a raw database (400 records) for the pre-processing stage.

E. Data Pre-processing

Pre-processing the data yields a structured dataset ready for use as input to the clustering model. This process involves two initial steps. First, reverse scoring is applied to negative statements (Q2, Q4, Q6, and Q9) to align the assessment's direction, ensuring that higher scores indicate a more positive assessment. Second, the aspect score or the average value of the relevant questions is calculated to form three main input features: GUI, Navigation, and Responsiveness.

These aspect scores for GUI, Navigation, and Responsiveness are calculated for each respondent *i* as the average of the grouped question items, as shown in Equations (1), (2), and (3). The score for each usability aspect is calculated as the average of the question items that comprise that aspect for each *i* respondent, after reverse-scoring negative items. Where the $Q1_i$ variable is the score of the *1*-th respondent on the *i*-th item, which has been adjusted through reverse scoring for negative statements (Q2, Q4, Q6, and Q9). The GUI, navigation, and responsiveness for each represent the average user-perception scores for interface quality, ease of navigation, and system responsiveness.

$$GUI_i = \frac{Q1_i + Q2_i + Q3_i + Q4_i}{4} \tag{1}$$

$$Navigation_i = \frac{Q5_i + Q6_i + Q7_i}{3} \tag{2}$$

$$Responsivenessf_i = \frac{Q8_i + Q9_i + Q10_i}{3} \tag{3}$$

The quality of the clustering results is evaluated using the Silhouette Coefficient, which is calculated for each data point *i*. The silhouette coefficient is computed using Equation (4). Where, the *a(i)* variable is the average distance between the i-th data point and all other data points in the same cluster (intra-cluster distance), The *b(i)* variable is the minimum average distance between the *i*-th data point and all data points in the nearest other cluster (nearest-cluster distance). The value of $s(i)$ variable falls within the range of $[-1,1]$, where a value approaching 1 indicates good cluster separation and high internal cohesion.

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (4)$$

The final stage of pre-processing is the normalization of all three features (GUI, Navigation, Responsiveness) using Z-score normalization (StandardScaler). This normalization makes each feature on a comparable scale, with values expressed relative to the mean and standard deviation of each aspect. This condition is important to prevent the dominance of any feature due to scale differences when applied to the X-Means Clustering algorithm, which is sensitive to distances in feature space.

F. Clustering Modelling and Analysis

The data analysis procedure centres on clustering using the X-Means Clustering Algorithm. The X-Means process is illustrated in Fig.2 [21]. This algorithm was chosen for its ability to automatically determine the optimal number of clusters (K-values), eliminating the need for manual K determination [22]. The initial parameters of the algorithm are set to a minimum *K* value and a maximum K. This modelling is intended to produce two Clusters in accordance with the research objectives. Although the X-Means algorithm automatically determines the optimal number of clusters, the final cluster configuration was selected based on the highest Silhouette Score value to ensure optimal separation and internal cohesion.

Fig.2. Flow of the X-Means Clustering Algorithm

DOI : https://doi.org/10.25139/inform.v11i1.11277

The Evaluation and post-modelling analysis cluster quality evaluation uses internal metrics, such as the Davies-Bouldin Index (DBI) and Silhouette Score, to assesshe quality of cluster separation and cohesiveness [23]. The model is stopped when there is no improvement in the evaluation metrics [24]. Centroid analysis and cluster profiles perform a Centroid Profile analysis (Fig.3 and Fig.4) to interpret the unique characteristics of each user group. This analysis serves as the basis for Equationting improvement recommendations and determining improvement priorities [25]. The entire process of calculation, algorithm modelling, and data visualization is carried out using the Python programming language in the Google Colab environment.

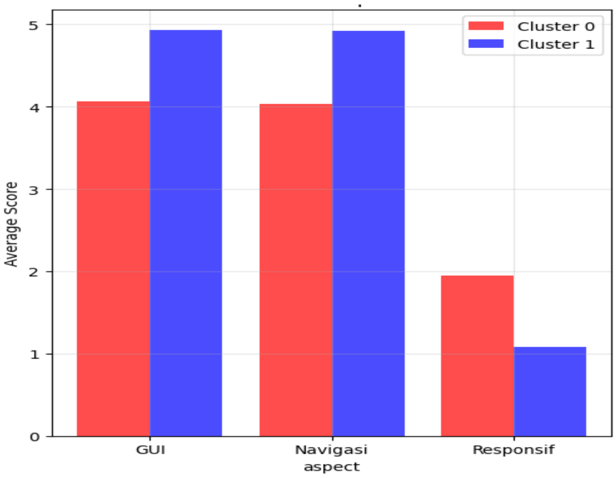

Fig. 3. Centroid Profile per Cluster

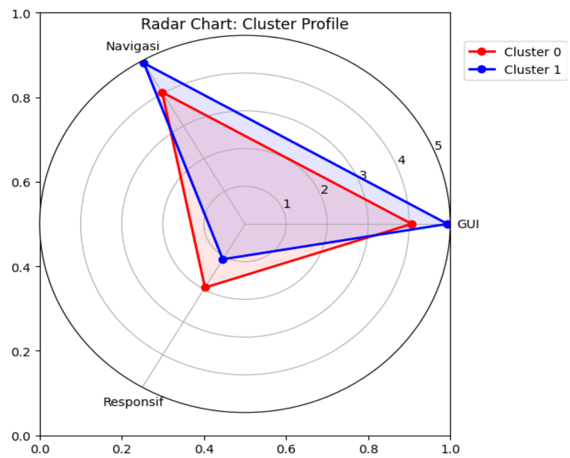

Fig. 4. Radar Chart Visualization

III. Result and Discussion

A. Test Instruments & Data Characteristics

The modified System Usability Scale (SUS) instrument was first tested with 80 participants to ensure its quality before use in the main data collection. The instrument consists of 10 statements compiled to measure three dimensions of user experience: the quality of the interface (GUI), ease and clarity of navigation, and system responsiveness. Validity testing was performed using Pearson's Product-Moment correlation, with results showing that all items had a correlation coefficient (r_{xy}) above 0.86, exceeding the minimum threshold of 0.30, and therefore all questionnaire items were declared valid. Internal reliability testing using Cronbach's Alpha yielded α = 0.974, indicating excellent internal consistency and confirming that the modified SUS instrument is highly reliable and feasible for measuring the usability of the Candy CBT application.

After the instrument was declared valid and reliable, the questionnaire was distributed to 400 active users of the CBT Candy application at *SMKS Panca Budi Medan*. Respondents came from four user categories: students, lecturers/teachers/stakeholders, and users of the application. All answers are recorded on a Likert scale of 1–5 with no missing values, so that all 400 records can be fully utilized in the pre-processing and clustering analysis stages. To maintain identity confidentiality, the respondent's name is replaced with a numerical code, so that data processing is carried out entirely on anonymized data. The list of modified System Usability Scale (SUS) questionnaire statements, which served as the research instrument, is presented in Table I. The ten statements are designed to capture user perceptions across the three main UX dimensions: the Graphical User Interface (Q1-Q4), Navigation (Q5-Q7), and Responsiveness (Q8-Q10). Statements Q2, Q4, Q6, and Q9 are negative statements (indicating dissatisfaction) and will require reverse scoring during data pre-processing to ensure that a consistently higher score reflects a more positive user experience.

TABLE I
MODIFIED SUS QUESTIONNAIRE STATEMENT

Label	Statement
Q1	The interface (GUI) of this CBT application is easy to understand.
Q2	The layout and icons on this CBT app are confusing.
Q3	Graphic design and colour support the comfort of use.
Q4	The GUI view made it difficult for me to use the app.
Q5	The navigation menu in this CBT application is clear and easy to use.
Q6	I often have trouble finding menus or features in this app.
Q7	The flow of the CBT application navigation feels consistent and logical.
Q8	This CBT app responded quickly to every command.
Q9	These CBT applications are slow or often do not respond well.
Q10	The CBT application runs stably without interference that inhibits response.

B. Data Pre-processing

Pre-processing the data yields a structured dataset ready for use as input to the clustering model. All 400 SUS questionnaire responses that have passed the validity and reliability tests are converted into three main features representing the CBT Candy app's usability dimensions: Graphical User Interface (GUI), Navigation, and Responsiveness. This conversion begins with aligning the assessment direction by reverse-scoring negative-statement items (Q2, Q4, Q6, and Q9), so that higher scores consistently reflect a more positive user experience. Each respondent is then represented by the average value of the aspect, obtained by grouping question items according to the measured dimensions. The data structure shifts from the level of the statement item to the level of the usability aspect, which is more concise and informative.

DOI : https://doi.org/10.25139/inform.v11i1.11277

To ensure consistency in score interpretation, reverse scoring was applied using a linear transformation, where the original Likert score was converted using the $reversed\,Score = 6 - Original\,Score$. For example, if a respondent rated item Q2 with a score of 2 (indicating disagreement), after reverse-scoring, the value becomes 4, which reflects a positive perception of usability. This transformation ensures that all questionnaire items follow the same direction, with higher values consistently representing better user experience before aspect aggregation and normalization.

The calculation of aspect scores using Equations (1) to (3) serves as a dimensionality reduction process, transforming the original 10-item SUS responses into three representative usability features. Each aspect score represents the mean of the items corresponding to respondent *i*. For instance, the GUI score for respondent *i* is calculated as the average of the items Q1–Q4 after reverse-scoring. If the adjusted scores for these items are 4, 5, 4, and 5, respectively, the resulting GUI score would be 4.50. This aggregation allows each respondent to be represented by a concise usability profile (GUI, Navigation, Responsiveness), which is more suitable for clustering analysis than individual questionnaire items.

The final stage of pre-processing is the normalization of all three features (GUI, Navigation, Responsiveness) using Z-score normalization. Z-Score Normalization was applied using the standard Equation $Z = (x - \mu) / \sigma$. Where x represents the original aspect score, μ is the mean value of the corresponding aspect across all respondents, and σ is the standard deviation. This normalization process transforms each feature to a standardized scale with a mean of 0 and unit variance.

The application of Z-Score Normalization is essential because X-Means clustering relies on distance-based calculations. Without normalization, aspects with higher average values, such as GUI and Navigation, would dominate the Euclidean distance computation, potentially biasing the clustering results. By standardizing all features, each usability aspect contributes proportionally to the clustering process, resulting in a more balanced and reliable cluster structure.

This normalization makes each feature on a comparable scale, with values expressed relative to the mean and standard deviation of each aspect. This condition is important to prevent the dominance of any feature due to scale differences when applied to the X-Means Clustering algorithm, which is sensitive to distances in feature space. Thus, the final dataset used in the modelling process is a 400 × 3 matrix, with each row representing a respondent's usability profile as standardized aspect scores. A summary of the features used, including the grouping of the original questionnaire items, is presented in Table II.

TABLE II
SUMMARY OF PRE-PROCESSING DATA RESULTS

Aspects and Feature Symbols	Forming Item Labels	Brief Definition
GUI	Q1, Q2, Q3, Q4	Average perception of display and interface quality
Navigation	Q5, Q6, Q7	Average perception of menu clarity and navigation flow
Responsiveness	Q8, Q9, Q10	Average perception of response speed and stability of CBT

The data pre-processing stage successfully converts the 400 raw SUS questionnaire responses into a structured dataset with three main features: GUI, Navigation, and Responsiveness. As shown in Table II, these features are derived from the average scores of specific item labels after reverse scoring (Q1-Q4 for GUI, Q5-Q7 for Navigation, and Q8-Q10 for Responsiveness), which have been adjusted to ensure consistency in the assessment direction. The final dataset is a 400×3 matrix of standardized aspect scores, ready for clustering.

C. Initial Usability Score Statistics

Preliminary statistics of usability scores calculated from 400 respondents show that the two main aspects, namely GUI and Navigation, are in the very good category, with average scores of around 4.53 and 4.51, respectively. In contrast, the Responsiveness aspect obtained an average score of only around 1.48 and was in the very low category. These results show that the majority of users rate the interface and navigation flow of the CBT application as very satisfactory. At the same time, response speed and system stability are the main weaknesses that most significantly affect the application's usability.

D. X-Means Clustering Results

Based on the research stages presented in Fig. 1, this section presents the results of the analysis, clustering model implementation, and testing phases, following the data collection and data pre-processing stages.

The main focus of this analysis is the application of the X-Means Clustering algorithm to questionnaire data obtained from 400 respondents. The clustering process was carried out in several sequential stages in accordance with the X-Means Clustering procedure, including data input preparation, initial cluster formation, cluster splitting evaluation, model selection, and final cluster validation. At the initial stage, a standardized dataset comprising three usability aspects (GUI, Navigation, and Responsiveness) was used as input for the clustering process. The X-Means algorithm begins by forming an initial partition using a minimum number of clusters, which then serves as the basis for further evaluation. In the next stage, each cluster is iteratively evaluated for potential splitting using statistical criteria, where the algorithm compares parent clusters and their candidate child clusters.

The X-Means algorithm automatically evaluates the data structure and determines the optimal number of clusters based on internal validation metrics. As a result of this evaluation process, the algorithm identified two optimal clusters ($K=2$) as the most suitable configuration for representing the usability patterns of CBT application users. The selection of these two clusters was further validated using the Silhouette Score, which measures cluster cohesion and separation. The highest

DOI : https://doi.org/10.25139/inform.v11i1.11277

Silhouette Score of approximately 0.78 was obtained with K=2, which is superior to alternative configurations (e.g., K=5, with a Silhouette Score of approximately 0.73). This result indicates that the CBT application's usability data structure is most stable and well-separated when partitioned into two user groups.

The final stage of the clustering process produces the distribution of respondents across the selected clusters. The results show a relatively balanced distribution: Cluster 1 comprises 215 respondents (53.8%) and Cluster 0 comprises 185 respondents (46.2%). This proportional distribution, as visualized in Fig.5, indicates that the clustering outcome is representative and suitable for comparative analysis across clusters. A summary of the clustering results, including centroid values and cluster profiles, is presented in Table III.

Based on Table III, both clusters exhibit high average scores for GUI and Navigation; however, the Responsiveness scores fall into the low category for all clusters. Cluster 1 shows higher GUI and Navigation scores than Cluster 0, but it records the lowest Responsiveness score. This pattern indicates that although users highly appreciate the interface design and navigation flow, system responsiveness remains the primary usability weakness across both clusters, regardless of their overall satisfaction level.

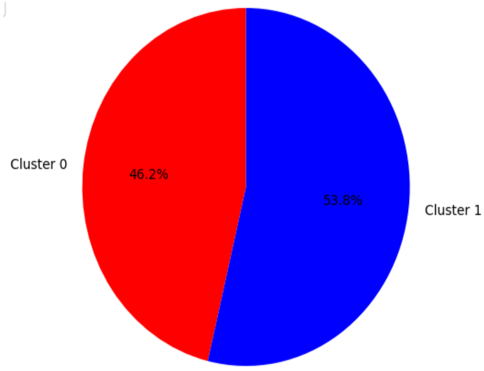

Fig.5. Respondent Distribution

To further validate the clustering structure, the separation of clusters in the feature space is visualised using Principal Component Analysis (PCA), as shown in Fig. 6. This visualisation represents the final stage of the clustering analysis. It illustrates the spatial distribution of Clusters 0 and 1 in reduced dimensionality. The clear separation between clusters, particularly along the first principal component, confirms that the X-Means Clustering process successfully identified distinct patterns of usability assessment among CBT application users.

TABLE III
X-MEANS CLUSTERING RESULTS (CENTROID DISTRIBUTION AND PROFILE)

Cluster	Number of Respondents	Percentage	Average GUI Score	Average Navigation Score	Average Responsiveness Score	Strength	Debilitation
1	215	53.8%	4.93	4.92	1.08	GUI, Navigation	Responsiveness
0	185	46.2%	4.06	4.03	1.95	GUI, Navigation	Responsiveness

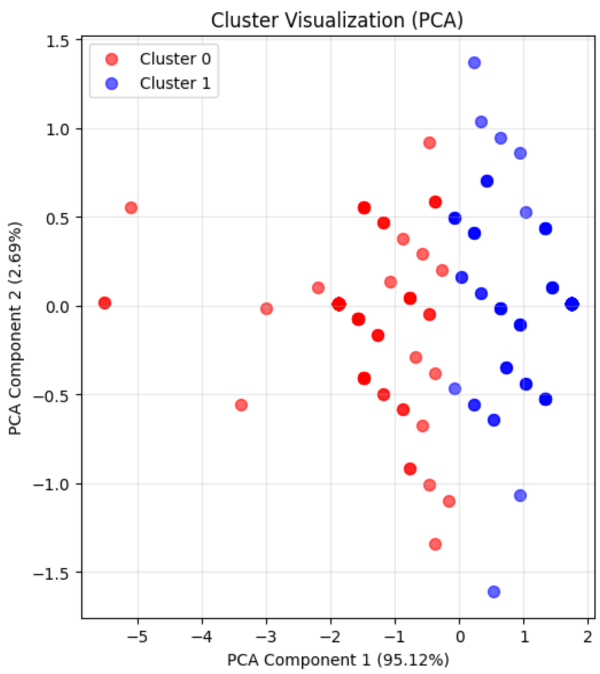

Fig. 6. Cluster (PCA)

E. *Centroid Profiles & Cluster Interpretations*

Centroid profile analysis based on Table III revealed distinct yet consistent patterns of satisfaction across both clusters, as shown in Fig.3 (bar chart) and Fig.4 (radar chart). Cluster 1, as the majority group, showed very high levels of satisfaction with the GUI (4.93) and navigation (4.92), indicating that the interface and the application menu flow are rated very well by users in this group. However, this cluster actually shows a very extreme level of dissatisfaction in the Responsiveness aspect, with a score of only 1.08. In contrast, Cluster 0 showed a more moderate pattern, with a GUI score of 4.06 and a navigation score of 4.03, while the Responsiveness score was 1.95, which, although higher than Cluster 1, remained in the low category. This pattern shows that the Responsiveness aspect is a major weakness and is consistently reported by both user groups. These differences in rating levels indicate that the higher the user's expectations of display quality and ease of navigation, the sharper their complaints about the speed and stability of the system's response. This is also reinforced by the radar chart in Fig.4, which shows that both clusters have large radii on the GUI and Navigation axes, while on the Responsiveness axis, both have much lower values, with Cluster 1 at the lowest. This visualization confirms that the Responsiveness dimension is the most dominant weak point in the user experience of CBT applications.

F. *Model Quality Evaluation*

The quality of the clustering results was assessed using Silhouette Analysis. This analysis uses measures of cluster separation and cohesion to assess the stability and quality of the

DOI : https://doi.org/10.25139/inform.v11i1.11277

final grouping. The Silhouette coefficient for each data object is calculated as a core step in this evaluation. This is defined by Equation (4).

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (4)$$

Equation (4) defines *s(i)* as the Silhouette coefficient for point *i*, where *a(i)* is the average distance from point *i* to all other points in the same cluster, and *b(i)* is the average distance from point *i* to the nearest cluster outside the cluster itself.

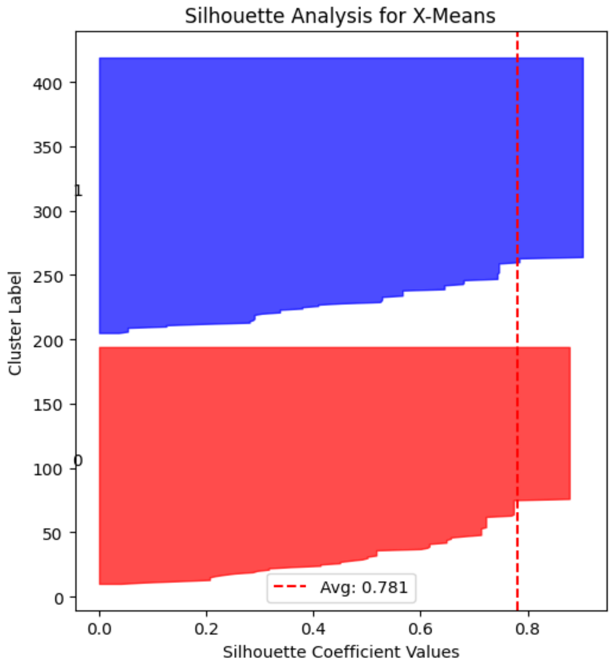

Fig.7. Silhouette Analysis

The results of the analysis are shown in Fig. 7. The average Silhouette Score for the two-cluster configuration was 0.781. This value is relatively high and close to 1, indicating that the clusters have a high level of cohesion, with data within each cluster relatively homogeneous and a clear separation between the formed clusters. In comparison, the configuration of the number of other clusters results in a lower Silhouette value, for example, about 0.73 when K=5, which is why it was not selected as the final model. Fig.7 shows that most Silhouette values for Cluster 0 and Cluster 1 members were above 0.5, confirming that the majority of respondents were assigned to clusters that matched the data structure. These results confirm that the two clusters are the most feasible configurations to represent variations in the perception of CBT application usability.

G. Improvement Priorities

The improvement priority is focused on the Responsiveness dimension, which has the lowest average value across the cluster. The priority score of each cluster is calculated using Equation (5). The priority score (*Prioritas$_c$*) is calculated by subtracting the cluster's average responsiveness score (*ResponsivenessScore$_c$*) from a base value of 10. Based on this calculation, Cluster 1 had a priority score of 8.92 (10 - 1.08), while Cluster 0 had a priority score of 8.05 (10 - 1.95). A summary of the results of this calculation is presented in Table IV.

$$Priority_c = 10 - Responsiveness\ Score_c \quad (5)$$

TABLE IV
PRIORITY OF IMPROVEMENT BASED ON RESPONSIVENESS ASPECTS

Cluster	Responsiveness Score	Priority Score	Priority Statement
1	1.08	8.92	Very high (top priority)
0	1.95	8.05	High (advanced priority)

Table IV shows that Cluster 1 falls into the very high priority category, while Cluster 0 is in the high priority category as a follow-up focus, indicating that both clusters require urgent attention to the Responsiveness dimension. Because Cluster 1 also has the largest number of respondents, which is 215 respondents or 53.8% of the total sample, the improvement of application responsiveness should be prioritized on the characteristics of users in this cluster. This group of users already shows a very high level of satisfaction with the GUI and Navigation aspects. Still, their user experience is severely hampered by slow response and system instability. These findings provide an empirical basis for developers to focus remedial resources on improving response speed and system stability, and on reducing technical glitches, before proceeding with further development of the display and navigation aspects.

H. Discussion

The results of this study show that the X-Means Clustering algorithm can determine and divide 400 respondents who use the CBT Candy application into two optimal clusters. These findings indicate that user experience (UX) cannot be interpreted as a single entity, but rather as a set of subgroups with significantly different patterns of satisfaction with functional and non-functional aspects. The formation of two user clusters characterized by differences across (GUI, Navigation, and Responsiveness aspects) has the impact that each cluster points to a unique point of weakness. The centroid profile of this cluster serves as an accurate diagnostic map: one group may appreciate navigation but is bothered by the system's responsiveness. In contrast, the other group has the opposite pattern. A direct consequence is that the improvement recommendations should not be generic but rather targeted to the priority needs of each cluster to maximize their impact.

The findings derived from the two resulting clusters methodologically distinguish this study from previous research. A study conducted by [9], which also employed X-Means Clustering on System Usability Scale (SUS) data, reported that the most stable clustering outcome consisted of two clusters based solely on the overall SUS score. Although the number of clusters obtained in the present study is the same, a fundamental difference lies in the analytical approach and the features used. This study addresses a research gap in prior work, which tended to generalize usability evaluations by relying exclusively on aggregated SUS scores and insufficiently exploring user experience (UX) dimensions in depth. This gap is addressed by

DOI : https://doi.org/10.25139/inform.v11i1.11277

modifying the clustering input features, with the clustering process using the average scores of three key UX aspects (GUI, Navigation, and Responsiveness) rather than the overall SUS score alone. This approach strengthens previous research by demonstrating that, even when X-Means produces the same number of clusters, the resulting cluster structures capture more specific and meaningful UX characteristics. Consequently, this study extends the theoretical understanding of clustering techniques in usability evaluation by highlighting that decomposing usability metrics enables a more granular, interpretable, and accurate diagnosis of system usability and user experience quality.

The main significance of the two-cluster model lies in its ability to identify the non-functional needs of the CBT Candy application in detail. Each cluster has a centroid profile that leads to specific improvement recommendations. For example, a group of users who are sensitive to responsiveness is demanding improvements in system performance, stability, and the ability to customize displays across multiple devices. In contrast, groups of users who are sensitive to Navigation demand a more intuitive and easy-to-understand flow design without the need for additional training. These results directly support the research objective of compiling specific and targeted improvement recommendations. As such, X-Means Clustering serves as both a data analysis tool and an efficient prioritization mechanism, allowing developers to allocate resources to the most impactful aspects for the most critical user groups.

This research theoretically contributes to the data mining and computer science literature by validating the aspect-enriched X-Means analysis model (GUI, Navigation, Responsiveness) as a new standard for more granular interpretation of SUS data. This result has a direct impact on *SMKS Pancabudi Medan*, as the resulting recommendations will be presented as a comprehensive list of non-functional needs. These improvements will ensure a more stable, user-friendly exam system, minimize technical disruptions, and make it easier for teachers to manage the question bank and monitor exam performance in real time. However, it should be noted as a limitation of the study that this study is entirely based on quantitative perception data (questionnaires). This reliance on self-reported perception data may affect the generalizability of the findings, as perceived usability does not always fully reflect actual user interaction behaviour. Therefore, the resulting cluster profiles should be interpreted as perceptual UX patterns rather than direct representations of real-time system performance. Therefore, the interpretation of cluster profiles has not been validated with actual user behaviour data or qualitative research, which is a suggested direction for future study development.

IV. CONCLUSION

This study presents an aspect-based usability evaluation of the Candy CBT application by integrating the System Usability Scale (SUS) with the X-Means clustering algorithm, by decomposing usability into three dimensions: GUI, Navigation, and Responsiveness. The analysis demonstrates that user experience assessment becomes more informative when it moves beyond a single aggregated SUS score. The X-Means algorithm successfully identified two optimal user clusters, indicating that aspect-level clustering provides a clearer and more interpretable segmentation of user perceptions.

The findings reveal a consistent pattern across both clusters: GUI and Navigation are rated very highly, whereas responsiveness is rated extremely low. This result indicates that system performance, response speed, and stability constitute the primary usability weakness of the CBT application, even among users who are otherwise satisfied with the interface design and navigation flow. Consequently, usability improvement efforts should prioritize non-functional requirements, particularly system responsiveness, rather than further enhancements to visual or navigational elements. Despite its contributions, this study has limitations. The analysis relies solely on quantitative perception data obtained from questionnaires, without validation through qualitative insights or actual user behaviour data. In addition, the clustering results were not compared with other clustering algorithms to assess robustness. Future research should incorporate behavioural logs, usability testing, or qualitative interviews, as well as comparative analyses with alternative clustering methods, to strengthen validation and enhance the generalizability of the proposed evaluation model. Such validation would enable future studies to examine the consistency between perceived usability clusters and actual user behaviour in real CBT use cases.

REFERENCES

[1] N. Romadhona, W. Fachrudin Yusuf, A. Yusuf, and A. Marzuki, “Penerapan Aplikasi Computer Based Test (Cbt) Sebagai Alternatif Penilaian Hasil Pembelajaran Pai Di Smk Syamsul Arifin Kraton Pasuruan,” no. 08, pp. 190–201, 2024.

[2] G. Daniel and A. Ikpe, "Computer-Based Testing (CBT): A Paradigm Shift in Examination Administration in Nigerian Post- Primary Schools," *International Journal of Modern Science and Research Technology*, vol. 3, no. 9, pp. 289–302, 2025, doi: 10.5281/zenodo.17207075.

[3] M. L. Hutahaean, E. Hadinata, and Y. Faradillah, “Aplikasi Computer Based Test (CBT) Berbasis Website Menggunakan Metode RAD Pada SMA Negeri 21 Medan,” *Jatilima : Jurnal Multimedia Dan Teknologi Informasi*, vol. 07, no. 01, pp. 1–10, 2025, doi: https://doi.org/10.54209/jatilima.v7i01.1085.

[4] F. D. Pratama, I. Zufria, and T. Triase, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 1, pp. 77–84, Jan. 2022, doi: 10.36341/rabit.v7i1.2217.

[5] M. Y. Kurniawan and T. Dirgahayu, “Evaluasi dan Rekomendasi Peningkatan User Interface dan User Experience pada Aplikasi Prime Video Mobile,” *Jurnal Kridatama Sains Dan Teknologi*, vol. 6, no. 02, pp. 453–468, 2024, doi: 10.53863/kst.v6i02.1221.

[6] M. R. Silalahi, L. M. Michelli, H. Umayasyah, D. A. Mu’adin, and B. Parga Zen, “Evaluasi Heuristik Dan System Usability Scale UI/UX pada Aplikasi ‘Makan Kuy,’” *Jurnal Ilmiah Media Sisfo*, vol. 18, no. 1, pp. 57–67, Apr. 2024, doi: 10.33998/mediasisfo.2024.18.1.1475.

[7] S. D. Rahmawati and B. Prasetiyo, "Application of Lean UX and System Usability Scale (SUS) Methods in Redesigning User Interface and User Experience on Adella Hospital Online Registration Website," *Journal of Advances in Information Systems and Technology*, vol. 6, no. 2, pp. 200–218, 2025, doi: 10.15294/jaist.v6i2.2430.

DOI : https://doi.org/10.25139/inform.v11i1.11277

[8] R. R. Khaerullah, N. Suarna, and O. Nurdiawan, "Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means," *Jurnal Informatika dan Teknologi Informasi*, vol. 1, no. 2, 2023, doi: 10.56854/jt.v1i2.135.

[9] Z. Manzis, R. Kurniawan, and Y. A. Wijaya, "Penerapan X-Means Clustering Terhadap Hasil Kuesioner Website Virtual Tour Stmik Ikmi Cirebon System Usability Scale (SUS)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2863–2869, 2024.

[10] M. N. Romadhoni and T. Dirgahayu, "Evaluasi dan Redesain UI/UX pada Aplikasi Web Young on Top," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 2390–2401, 2024, doi: 10.35870/jimik.v5i3.909.

[11] V. A. Subarjah and Ari Purno Wahyu, "Analysis and Design of User Interface and User Experience of Regional Tax Enterprise Resources Planning System with Design Thinking Method," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, pp. 96–106, 2022, doi: 10.25139/inform.v7i2.4729.

[12] T. Rahmawati, "Analisis perbandingan silhouette coefficient dan metode elbow pada pengelompokkan provinsi di indonesia berdasarkan indikator ipm dengan k-medoids 1,2,3," vol. 13, pp. 13–24, 2024, doi: 10.14710/j.gauss.13.1.13-24.

[13] H. A. Ulvi and M. Ikhsan, "Comparison of K-Means and K-Medoids Clustering Algorithms for Export and Import Grouping of Goods in Indonesia," *sinkron*, vol. 8, no. 3, pp. 1671–1685, Jul. 2024, doi: 10.33395/sinkron.v8i3.13815.

[14] A. Khamaj and A. M. Ali, "Adapting user experience with reinforcement learning: Personalizing interfaces based on user behavior analysis in real-time," *Alexandria Engineering Journal*, vol. 95, pp. 164–173, May 2024, doi: 10.1016/j.aej.2024.03.045.

[15] J. R. Leite Filho, T. A. Coleti, and M. Morandini, "Evaluating chatbot user experience (UX) through electroencephalography measures: A systematic literature review," Mar. 2024, *Elsevier B.V.* doi: 10.1016/j.chbr.2024.100381.

[16] N. Septiana and M. Rohmadi, "Persepsi Mahasiswa Terhadap Aplikasi CBT Android Berbasis Exambrower Sebagai Alat Evaluasi Paper Less," *JRIP: Jurnal Riset dan Inovasi Pembelajaran*, vol. 4, no. 1, pp. 448–464, 2024.

[17] R. Rahim, "Pelatihan Sistem Informasi Penerimaan Siswa Baru Di Smk Pancabudi Medan," *Jurnal PKM Bangsa (Jurmas Bangsa)*, vol. 1, no. 2, pp. 95–100, 2023, doi: 10.62357/jpb.v1i2.175.

[18] D. S. Atanggae, A. Angelaus Ete, and M. Dhema, "Keefektifan Ujian Berbasis Komputer Menggunakan Aplikasi CANDY CBT Di SMA Muhammadiyah Maumere," *Jurnal PTI (Jurnal Pendidikan Teknologi Informasi)*, vol. 11, no. 2, pp. 50–55, 2024, doi: 10.35134/jpti.v11i2.219.

[19] D. Ayu and B. Utami, "Perancangan Sistem Login Pada Aplikasi Berbasis GUI Menggunakan QTDesigner Python," vol. 4, no. 2, pp. 92–100, 2021, doi: https://doi.org/10.30873/simada.v4i2.2961.

[20] A. A. Suhendra, G. Agung, A. Putri, G. Made, and A. Sasmita, "Evaluasi Usability User Interface Website Menggunakan Metode Usability Testing Berbasis ISO 9241-11 (Studi Kasus PT.X)," *JITTER-Jurnal Ilmiah Teknologi dan Komputer*, vol. 2, no. 3, pp. 1–9, 2021.

[21] E. Timothy Napitupulu and F. Aldi Jous Nainggolan, "Perbandingan Algoritma K-Means Dan X-Means Untuk Mengelompokkan Minat Kejuruan Siswa Baru Pada Smk Multikarya Medan Menggunakan Metode Clustering," *Jurnal TEKINKOM*, vol. 6, no. 2, p. 2023, doi: 10.37600/tekinkom.v6i2.933.

[22] E. Fitriana, Windarsyah, and Kamarudin, "Clustering Kos Dengan Algoritma K-Means Untuk Rekomendasi Tempat Berdasarkan Harga dan Fasilitas," *Jurnal Media Informatika [JUMIN]*, vol. 6, no. 3, pp. 1990–1995, 2025.

[23] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2697.

[24] B. Yusuf, R. Mahara, H. Ahmadian, S. Wahyuni, P. Studi Teknologi Informasi, and P. Sudi Pendidikan Teknologi Informasi, "Analisis Clustering Penduduk Miskin Di Provinsi Aceh Menggunakan Algoritma K-Means Dan X-Means," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.

[25] R. Hariyanto and M. Z. Sarwani, "Optimizing K-Measn Algorithm Using Particle Swarm Optimization To Group Student Learning Processes," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 6, no. 1, pp. 65–68, 2021, doi: 10.25139/inform.v6i1.3459.