Comparative Analysis of Clustering Approaches in Assessing ChatGPT User Behavior

Dedy Setiawan*[1], Daniel Arsa[2], Lucky Enggrani Fitri[3], Farah Fadhila Putri Zahardy[4]
[1,2,4]Program Studi Sistem Informasi, FST Universitas Jambi, Jambi
[3]Jurusan Ilmu Ekonomi, FEB Universitas Jambi, Jambi
e-mail: ***[1]dedy_setiawan@unja.ac.id**, [2]danielarsa@unja.ac.id , [3]lucky_fitri@unja.ac.id, [4]farahfadhilapz@gmail.com

Abstract

ChatGPT is an artificial intelligence technology that is widely used and discussed. The technology invites mixed responses from various parties, mainly because of the benefits and risks of its use in multiple fields. Jambi University students also feel the influence of ChatGPT's presence in education. To determine the behavior of Jambi University students in using ChatGPT, four UTAUT variables were used, namely Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), and Facilitating Condition (FC) as independent variables in measuring the behavior of using ChatGPT. According to UTAUT, these four variables positively influence the actual behavior of technology use. This study used K-Means and K-Medoids Clustering to group Jambi University students based on ChatGPT usage behavior. Based on the Silhouette Score calculation, each method's optimal number of clusters is 2. K-Means is considered more optimal in forming 2 clusters because it obtained a Silhouette Score of 0.2123864, higher than K-Medoids, which is 0.1766865.

Keywords— ChatGPT, K-Means, K-Medoids, UTAUT, Silhouette Score

1. INTRODUCTION

Since new-generation information technologies like blockchain, IoT, and AI (Artificial Intelligence) are being used widely, technological and industrial change is happening more quickly [1]. On November 30, 2022, OpenAI launched a model that can interact with users via conversation, namely ChatGPT. When in conversational mode, ChatGPT can answer additional queries, acknowledge errors, refute presumptions, and deny inappropriate requests [2]. Five days after its launch, ChatGPT gathered 1 million users and became the fastest online service to achieve this [3]. ChatGPT is attracting much attention from government, industry, academia, and the public. ChatGPT's ability to produce human-like writing has given rise to various opinions. ChatGPT and other language models have been described as a double-edged sword. Despite the challenges in using AI-based technologies like ChatGPT, harnessing the power of technology for decision support has great potential [4].

This study seeks to examine the behavior of Jambi University students in utilizing ChatGPT. This generative AI model has ignited extensive discourse over the significance of AI technology, its educational applications, and its effects on student development, learning, assessment, and certification. While several educators and practitioners assert that ChatGPT can facilitate the learning and development process and advocate for its regulated use, others express apprehension that this technology may undermine the essential objectives of education [5]. The exposure and impact of ChatGPT inevitably influence students at Jambi University. Consequently, the study is required to ascertain and gather data regarding their behavior in utilizing ChatGPT. In a survey on technology adoption and diffusion, the Unified Theory of Acceptance and Use of Technology (UTAUT) is frequently employed as a theoretical framework to examine user intents and behaviors. Prior research indicates that variables including PE, EE, SI, and FC positively influence technology usage behavior [6]–[8]. This study employs the framework to examine data gathered from students and categorize users according to their behavior. This segmentation uses the clustering method in data mining to gain a deeper understanding of the characteristics and behaviors of ChatGPT users.

Clustering is a technique for categorizing items according to specific attributes and the degree of similarity between them.[9]. K-means and K-medoids are two prevalent methods utilized in clustering. K-Means is an algorithm that categorizes data into K clusters to minimize the overall distance within each cluster. This method is recognized for its computational efficiency and capacity to manage large-scale data with rapid processing times [10]. Nonetheless, K-Means has a vulnerability to outliers, as high-value items can considerably influence the data distribution pattern. K-Medoids serve as a solution to address this limitation by utilizing a single object as the cluster center, hence enhancing robustness against uneven data distribution [11]. The primary distinction between these two algorithms lies in the approach employed to ascertain the cluster centroids; K-Means utilizes the mean value, whereas K-Medoids employ a singular item to represent each cluster [12]. Despite the typical application of both algorithms, research evaluating the efficacy of K-Means and K-Medoids, mainly for user behavior clustering in technologies like ChatGPT, still needs to be explored. Prior research indicates variability in outcomes, with K-Means demonstrating superior performance in certain instances [13], [14], while K-Medoids yield more favorable results in others [15], [16]. The selection of an appropriate algorithm is very contextual and contingent upon the qualities of the data under analysis.

This study aims to address the gap in the literature about ChatGPT user behavior by merging the UTAUT framework with clustering approaches. This study also seeks to furnish additional empirical information regarding the efficacy of the K-Means and K-Medoids algorithms in clustering user behavior. The findings are anticipated to enhance academic literature and offer practical insights to aid in the formulation of policies and strategies for the implementation of AI technology in education.

2. RESEARCH METHODS

This research was executed through systematic phases illustrated in the Research Stages presented in Figure 1. The initial phase commenced with a literature review to comprehend the primary challenges associated with the application of ChatGPT in education alongside pertinent theoretical frameworks, including the UTAUT model, which assesses elements such as PE, EE, SI, and FC. The subsequent phase involves data collection, which is executed via an online survey directed at students of Jambi University through the distribution of questionnaires utilizing Google Forms. The acquired data encompasses details regarding ChatGPT usage trends, which are subsequently analyzed. During the data pre-processing phase, data cleaning was performed to address missing or inaccurate values, normalization was performed to standardize the scale, and validation was performed to ensure data integrity.

The processed data is utilized in a data processor, where the Silhouette Score ascertains the optimal cluster count. Subsequently, the K-Means and K-Medoids algorithms are employed to cluster student data according to ChatGPT usage patterns. K-Means was selected for its efficiency in managing large-scale data and its capacity to generate compact, well-defined clusters. Nonetheless, due to K-Means' susceptibility to outliers that may influence clustering outcomes, the K-Medoids technique is employed as an alternative. K-medoids exhibit more robustness to unevenly distributed data as they utilize a single item as the cluster centroid, leading to enhanced stability in clustering outcomes. The integration of these two algorithms enables the study to assess the efficacy of each method and guarantees that the clustering outcomes possess excellent accuracy and relevance.

During the validation phase, the clustering outcomes of K-Means and K-Medoids were evaluated by metrics including cluster centroids, visual representations of clustering findings, and Silhouette Score values. This procedure guarantees that the employed clustering method yields optimal outcomes. Ultimately, the analysis of the results categorized student data according to

ChatGPT usage patterns, yielding comprehensive insights into user behavior and substantial contributions to the research and practice of AI technology management in education. This investigation concluded following a thorough analysis of all stages.

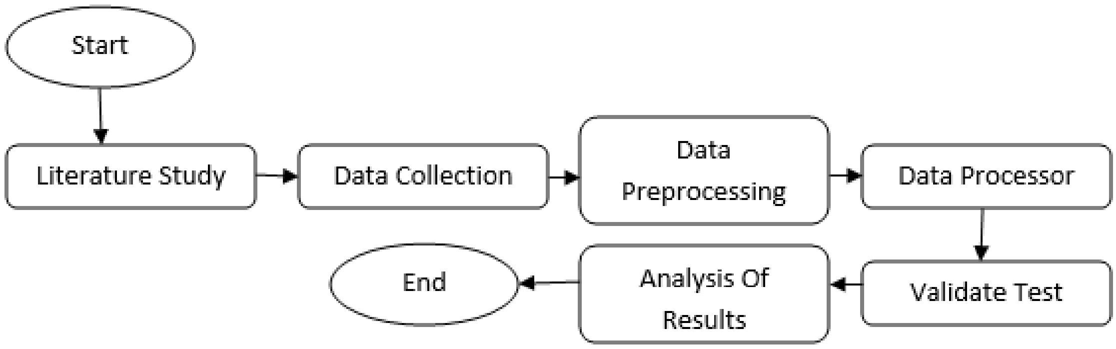


Figure 1 Research Stages

2.1. Variable Components

This study used four independent variables from the UTAUT model to examine the usage trends of ChatGPT among students at Jambi University. The four factors are PE, EE, SI, and FC. According to [17], PE is defined as the degree to which an individual believes that technology can enhance task completion efficacy, exemplified by students' perceptions of ChatGPT's assistance for their academic endeavors. EE assesses the perceived simplicity of utilizing technology, encompassing student evaluations of the interface's user-friendliness and the process of engaging with ChatGPT. Moreover, SI denotes the conviction that significant individuals, such as relatives or friends, deem it essential to utilize a specific technology. This variable indicates the degree to which environmental social pressure affects students' decisions about the use of ChatGPT. FC refers to the availability of sufficient resources and support, such as access to hardware, internet connectivity, or usage manuals, that affect students' capacity to utilize the technology.

The UTAUT framework posits that PE, EE, and SI directly affect user intention to utilize technology [18]–[20], while the intention to use, in conjunction with FC, dictates actual usage behavior [21]. This research utilizes four variables to elucidate the factors influencing Jambi University students' adoption of ChatGPT, thereby offering a thorough understanding of AI technology usage trends in educational settings.

2.2. Preprocessing Data

Preprocessing aims to convert data into a format that is easier to understand and increase the efficiency of the data mining process as needed. Preprocessing also contributes to achieving more precise results, reduces computing time for large datasets, and produces a denser data representation without losing important information. In this research, preprocessing was done by normalizing the data using Z-Score Normalization.

Z-Score, The process of normalization, scales the data values by using the average and standard deviation of each feature attribute. This normalization technique works well to lessen the influence of outliers. Equation (1) is used to calculate Z-Score Normalization [22]:

$$Z=(X-\mu)/\sigma \qquad (1)$$

With :

Z = Z-Score
X = initial value
μ = average of a dataset
σ = standard deviation.

2.3. K-Means

One generic technique for organizing data according to comparable attributes is K-Means. The K-Means algorithm divides data into k clusters, each of which has been initialized beforehand [23]. K-Means enables labeling, summarization, and sometimes a deeper understanding of hidden data dynamics [24].

The steps in the K-Means algorithm are as follows [10], [25], [26]:

1. Determine the data set's number of clusters (k).
2. Select centroid. In the initial stage, the centroid value is chosen randomly using a formula to set the initial K-Means target. Equation (2) calculates the target data or distance between groups, the initial center point for iteration 0 in the K-Means algorithm.

$$Initial\ Target = \frac{Number\ Of\ Data\ Points}{Number\ Of\ Classes + 1} \tag{2}$$

Information:

Number of Data Points = Total data to be used.

Number of classes = Predefined clusters, including categories such as very high, high, normal, low, and very low. An average formula is used during iteration, and calculations are performed to find the average value using Equation (3) below.

$$V_{ij} = \frac{\sum_{k=1}^{N_i} X_{kj}}{N_i} \tag{3}$$

Information:

V_{ij} = The centroid value of cluster i in dimension j.

N_i = Quantity of data points in cluster i.

X_{kj} = The k-th data value in dimension j.

$\sum_{k=1}^{N_i} X_{kj}$ = The aggregate of all values of dimension j for all data inside cluster i.

3. Determine the shortest path to the centroid in each record. Equation (4) is used to compute this centroid distance using the Euclidean Distance:

$$D_e(x_i , \mu_k) = \sqrt{\sum_{j=1}^{m} \left(x_{ij-\mu_{kj}}\right)}^{\,2} \tag{4}$$

Information:

$D_e(x_i , \mu_k)$ = Euclidean distance between data point xi and cluster centroid k

Xi = Data point xi with coordinates in the m-dimensional space.

μ_k = The centroid of cluster k possesses coordinates in dimension m.

Xij = The data values xi in dimension j.

μk = The centroid values μk in dimension j.

m = The quantity of dimensions inside the data.

2.4. K-Medoids

The K-Means method has been modified to become the K-Medoids algorithm. This approach chooses a representative item, known as a Medoid, for each cluster at each iteration rather than averaging the items in each cluster. The use of rules has two benefits: Medoids help describe clusters, and the K-Medoids technique determines the distance from the distance matrix, eliminating the need for repetitive distance calculations at each iteration [27].

K-Medoids, sometimes referred to as PAM (Partitioning Around Medoids), is a technique that divides data with n items into k clusters, where k is not greater than n. Therefore, the goal is to find k such objects. Grouping is done based on the similarity between objects, which is measured using a distance measure [28].

The following are the steps for completing K-Medoids [11], [29] :

1. Start by initializing the cluster centers as many as the number of clusters (k).
2. Find the original medoid. Equation (5) is used to use the Euclidean Distance equation to arrange each data or object into the nearest cluster.

$$D_e(x_i\,, m_k) = \sqrt{\textstyle\sum_{j=1}^{m}\left(x_{ij-m_{kj}}\right)} \quad (5)$$

With :

$D_e(x_i\,, m_k)$ =The Euclidean distance between object x_i and medoid m_k.
X_{ij} = The coordinate value of item x_i in dimension j.
m_{kj} = Medoid coordinate value m_k in dimension j.
n = Quantity of dimensions.

3. Choose a new medoid candidate at random from each cluster.
4. Determine the separation between the new medoid candidate and every object in each cluster.
5. Subtract the new total distance from the old total distance to determine the total difference (S). To create a new set of items as medoids, swap the objects with the cluster data if S is less than 0.
6. To achieve the correct cluster and cluster members, repeat steps 3 through 5 until the medoid remains unchanged.

2.5. Silhouette Score

Cluster evaluation using the Silhouette method was conducted to assess the extent to which the data was successfully grouped based on the predetermined clustering model [30]. By measuring the quality of the clusters created using the average silhouette value approach, the silhouette score is a technique for figuring out how many clusters there are. The better the silhouette value, the higher the average value [31]. Cluster quality, measured through the Silhouette value, can be seen from the extent to which objects in one cluster approach each other and the extent to which objects in different clusters approach each other. The following procedures are used to determine the silhouette coefficient value [32]:

1. Using Equation (6), determine the average distance between data i and every other data in the same cluster.

$$a\,(i) = \frac{1}{|C\,|-1}\textstyle\sum_{j\,\in\,C,j\neq i} d(i,j) \quad (6)$$

Where :
a (i) = The average distance between data point i and all other data points in cluster C.
|C| = represents the aggregate quantity of data within cluster C.
d (i,j) = denotes the distance between data point i and data point j.

2. Employing Equation (7) to compute the mean distance of object i to all objects in alternative clusters.

$$d\,(i, A') = \frac{1}{|A'\,|}\textstyle\sum_{j\,\epsilon\,A'} d\,(i,j) \quad (7)$$

Where :
d (i,A’) = Mean distance between data point i and all other data points in cluster A′.
|A’| = Quantity of elements (data) in cluster A′.
d(i,j) = Distance between data point i and data point j

.

$\sum_{j \in A'} d\ (i, j)$ = Calculation of the total distances between data point i and each data point j inside cluster A′.

3. Utilizing Equation (8), determine the value of the silhouette coefficient.

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (8)$$

Where :

s(i) = silhouette data i's value.

3. RESULT AND DISCUSSION

3.1. Data collection

[1]To comprehend how students use ChatGPT, four independent variables are required: PE, EE, SI, and FC. Technology use habit is positively impacted by these four factors. This behavior is related to the motivation to use ChatGPT again. The following are details of the variables and indicators used in this research.

Table 1. Variable Components

Variable	Dimensi
Performance Expectancy	Perceived ease of management Speed in doing Performance gains Motivation
Effort Expectancy	Ease of interaction Complexity Perceived ease of use Ease of learning
Social Influence	Family factors Friend factors Social factors Influential People
Facilitating Condition	Conditions that facilitate Knowledge Suitability Widely accepted

Based on Table 1, a questionnaire consisting of 16 statements representing each variable indicator was prepared. The measurement scale in the questionnaire is an ordinal scale with five levels as follows:

Table 2 Likert Scale

Scale	Information
1	Strongly Disagree
2	Don't agree
3	Agree
4	Strongly agree

Table 2 presents an ordinal scale to illustrate the hierarchy of assessment for a statement according to the degree of respondent agreement. This scale comprises four categories: 1 (Strongly Disagree), 2 (Disagree), 3 (Agree), and 4 (Strongly Agree). Each category delineates the hierarchy of preference or intensity of the respondent's attitude towards a statement, ranging from the lowest degree of disagreement to the highest degree of agreement. In an ordinal scale, the numerical value signifies the relative position or ordering of each category in the measurement rather than denoting equal intervals between levels. Next, the questionnaire was distributed online to Jambi University students. Data was obtained from 400 respondents who were Jambi University students who had used ChatGPT.

1

3.2. Data Preprocessing

The data processing procedure in this study commences with the acquisition of raw data using an online form. The received data has undergone preliminary verification to confirm compliance with quality standards, precisely the absence of missing information, consistency, and adherence to the anticipated response format. This phase is crucial to guarantee the authenticity and validity of the data before its use in subsequent research.

Upon validation of the raw data, a preprocessing phase is conducted to ready the data for comprehensive analysis. A crucial phase in preprocessing is data normalization, which seeks to standardize the value scale of each variable to facilitate effective data comparison. The normalization procedure employs the Z-Score approach, wherein each data value is computed based on its deviation from the mean and divided by the standard deviation. This normalization yields data with a mean of zero and a standard deviation of one, removing the impact of scale disparities among variables.

The normalization procedure is executed with the "scale()" function included in the CRAN package within the RStudio program. The NORM_P1 column is derived from the normalization of the P1 column. Figure 1 illustrates the results of the normalized data. This stage guarantees that the data is prepared for subsequent analytical procedures, including the implementation of clustering algorithms.

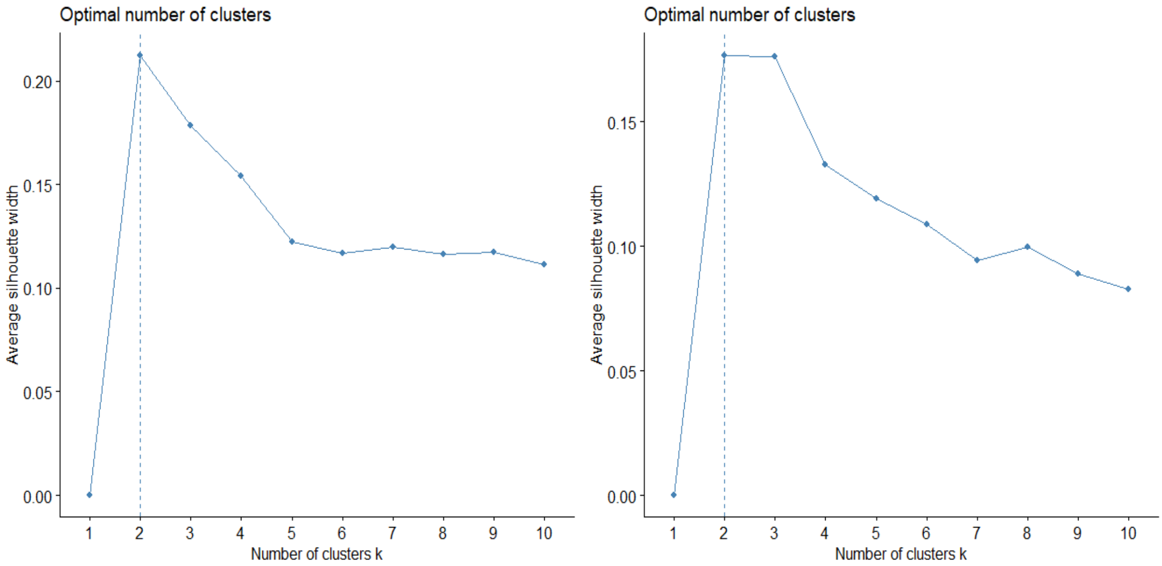
	RES	P1	NORM_P1
1	RES1	4	0.877820698743772
2	RES2	4	0.877820698743772
3	RES3	3	-0.432359448634992
4	RES4	4	0.877820698743772
5	RES5	4	0.877820698743772

Figure 2. Data Standardization

Normalization renders the data more uniform and pertinent for distance-based analytical methods, such as clustering. This phase is crucial for enhancing the precision and efficacy of the analytical methodology employed in the research.

3.3. Calculating the Ideal Number of Clusters

The Silhouette score computation yielded the ideal number of clusters for this study. An object's silhouette score indicates how well it fits within a cluster. The average Silhouette value in the dataset can be used to calculate the ideal number of clusters. The dataset's number of clusters is more perfect when the Silhouette value is more significant.



(a) *(b)*

Figure 3 Silhouette score

Figure 3(a) illustrates the Silhouette value derived from the number of clusters utilized in the K-Means algorithm, whereas Figure 3(b) presents the corresponding graph for the K-Medoids technique. The second graph depicts the correlation between the quantity of clusters and the clustering quality, as assessed by the Silhouette value. The K-Means graph (Figure 3a) demonstrates that the Silhouette value peaks at two clusters, signifying that partitioning the data into two clusters yields an optimal clustering configuration regarding internal cohesion and inter-cluster separability. The K-Medoids graph (Figure 3b) similarly demonstrates that a cluster count of 2 yields the highest Silhouette value, signifying that partitioning into two clusters is best for this technique. The elevated Silhouette values in both techniques signify that partitioning the data into two clusters yields substantial cohesion within the clusters (indicating that data points inside a cluster are closely grouped) and distinct separation between the clusters (demonstrating that different clusters are well delineated). Consequently, both methods determine that the best number of clusters is two, signifying the most efficient and representative data partitioning.

3.4. Clustering

Next, K-means and K-medoids are used to perform clustering. The sixteen statements that were prepared were subjected to clustering. R programming is used to carry out the clustering procedure. The "means ()" function in the "stats" package, which is accessible from the beginning in Rstudio, is used for clustering using the K-Means approach. Two clusters were produced as a result of this approach; Cluster 1 had 255 respondents, while Cluster 2 had 145. Meanwhile, K-Medoids clustering produces 2 clusters of 123 and 277 cluster members. The function used is "pam()" in the "cluster" package in Rstudio.

The following is the central point of the cluster formed:

Table 3 Cluster Center

	Cluster 1		Cluster 2	
	K-Means	*K-Medoids*	*K-Means*	*K-Medoids*
P1	3.67	4	2.72	3
P2	3.74	4	2.76	3
P3	3.40	4	2.48	3
P4	3.29	4	2.34	3
P5	3.25	4	2.54	3
P6	3.39	4	2.54	3
P7	3.82	4	2.93	3
P8	3.64	4	2.74	3

P9	1.76	2	1.59	2
P10	3.13	4	2.60	3
P11	1.93	2	1.70	2
P12	2.57	3	1.97	2
P13	3.51	4	2.58	3
P14	3.62	4	2.70	3
P15	3.43	4	2.50	3
P16	3.45	4	2.58	3
Average	3.23	3.69	2.45	2.81
Number of Members	255	123	145	277
Medoids	-	RES16	-	RES293

Table 3 presents the clustering outcomes derived from the K-Means and K-Medoids algorithms, each resulting in the formation of two clusters. In Cluster 1, K-Means exhibits an average value of 3.23 with 255 members. In contrast, K-Medoids demonstrates an average value of 3.69 with just 123 individuals, suggesting that K-Medoids allocates fewer members to Cluster 1. Conversely, Cluster 2 in K-Means has an average of 2.45 with 145 members, whereas K-Medoids demonstrates an average of 2.81 with a more significant membership of 277. The disparity in member count illustrates how the two algorithms delineate clusters, with K-Medoids typically aggregating a greater volume of data into Cluster 2. Furthermore, K-Medoids identify medoids that signify the central locations of the clusters, specifically RES16 for Cluster 1 and RES293 for Cluster 2, whereas K-Means employ centroids derived from the average of the cluster constituents. Both algorithms generate two clusters; however, they differ in the selection of cluster members and mean values, with K-medoids emphasizing the representation of cluster centroids via medoids.

This study considers differences in cluster center points significant if the difference distance is ≥1. This considers the type of scale used, namely a Likert scale with a distance between answer choices of 1. A difference distance of ≥1 indicates a change in level on the scale, so it is considered a significant difference. Table 3 allows for the following deductions to be made:

- The averages of clusters 1 and 2 in the K-Means and K-Medoids algorithms do not significantly differ from one another, according to the cluster average values. But compared to K-Means, the average cluster value in K-Medoids is higher.
- Overall, the cluster center points from K-Means and K-Medoids clustering do not differ significantly, with the K-Medoids center point being higher than the K-Means center point.

An illustration of the clusters created with K-Means and K-Medoids may be found below.

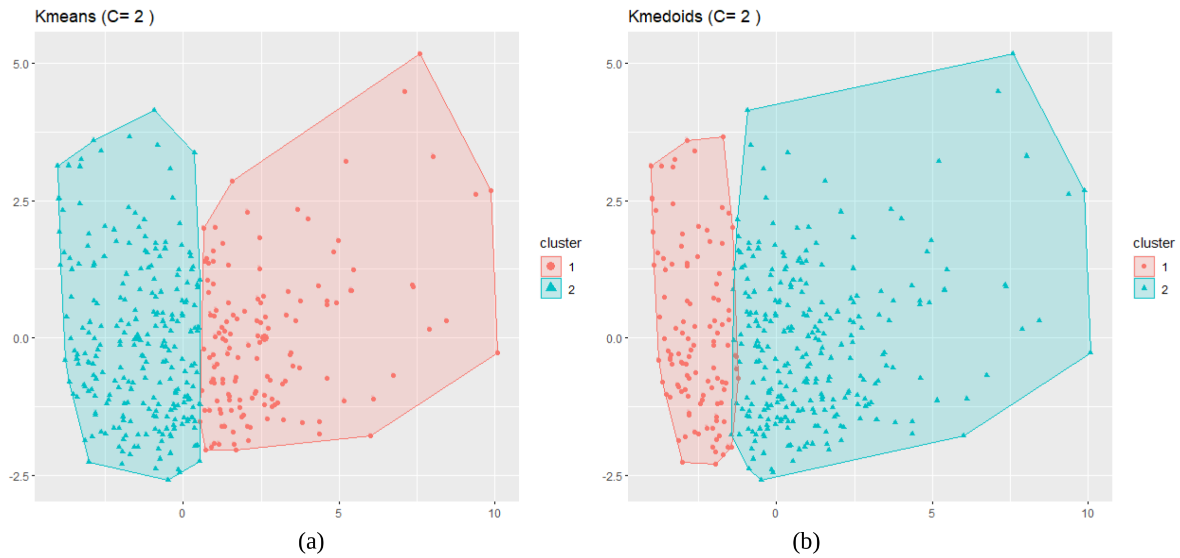


Figure 4 Visualization of frequent clusters k=2

.

Picture(a) shows a thin slice between cluster 1 and cluster 2 formed in clustering with K-Means. Meanwhile, clustering with K-medoids is shown. Picture(b), the formed wedge between cluster 1 and cluster 2 can be seen more clearly.

This research used the Silhouette Score to determine the best clustering method. Silhouette values range from -1 to 1. Positive values indicate that the point is closer to its cluster than neighboring clusters. This indicates that the quality of the clusters created increases with the Silhouette value. The outcomes of calculating the Silhouette Score k=2 are as follows:

Table 4 Silhouette Score

Silhouette Score	*K-Means*	*K-Medoids*
k=2	0.2123864	0.1766865

Table 4 shows that the Silhouette value obtained from grouping with K-Means is more significant than with K-Medoids. Therefore, it can be concluded that in this instance, clustering using K-Means is superior to using K-Medoids.

3.5. Discussion of Results

The analytical results indicated that the K-Means algorithm yielded superior clustering compared to K-Medoids in this study. Consequently, the K-Means clustering outcomes were employed to categorize Jambi University student data according to their ChatGPT usage behavior. The assessment of student behavior relies on four independent variables: PE (V1), EE (V2), SI (V3), and FC (V4), selected for their substantial impact on user behavior regarding technology adoption. The favorable experience derived from these elements is anticipated to enhance users' desire and opportunities to persist in utilizing the technology.

At this stage, the value for each variable is derived by computing the mean of the respondent's responses to the four affirmative statements associated with each variable, utilizing a specified Likert scale. The calculated average is subsequently rounded to yield a score between 1 and 4. The technique of ascertaining the value of this variable involves rounding the mean of the responses to each statement that represents the variable in question. P1-P4 denote the PE variable (V1), P5-P8 signify the EE variable (V2), P9-P12 indicate the SI variable (V3), and P13-P16 reflect the FC variable (V4).

Table 5 Determining Variable Values

RES	**P1**	**P2**	**P3**	**P4**	**V1**
RES1	4	4	3	3	**4**
RES2	4	4	3	3	**4**
RES3	3	3	3	3	**3**
RES4	4	4	3	3	**4**
RES5	4	4	3	3	**4**
RES6	1	2	2	2	**2**
RES7	3	4	3	3	**3**
RES8	4	4	4	3	**4**

The variable value is obtained by rounding the average value of the statement that represents the variable, and the cluster center based on the variable is as follows.

Table 6 Cluster Variable Value

Cluster	**V1**	**V2**	**V3**	**V4**	**Average (16 statements)**
C1	4	4	2	3	3.225
C2	3	3	2	2	2.45387925
Average (400 respondents)	3.28	3.36	2.36	3.28	

According to the analytical results presented in Table 6, two clusters exhibit distinct characteristics of ChatGPT users. Cluster C1 has elevated scores in PE (V1) and EE (V2), both rated at 4, signifying that participants in this cluster possess substantial anticipations regarding the advantages and user-friendliness of ChatGPT. The low SI (V3) score (2) means that social influence is not a key factor in their decision to adopt this technology. Furthermore, the FC (V4) score of 3 suggests that the supportive elements for utilizing ChatGPT are sufficient; however, there remains potential for enhancement. Cluster C1, with an average score of 3.225, predominantly represents users who exhibit optimism and satisfaction regarding their experience with ChatGPT, exhibiting minimal influence from external limitations. Conversely, Cluster C2 exhibits a modest score, with PE (V1) and EE (V2) both rated at 3, indicating diminished expectations regarding the technology's performance and usability. The diminished values of SI (V3) and FE (V4) (2) suggest that users within this cluster are minimally affected by social factors and perceive the environment supporting the utilization of this technology as suboptimal. Cluster C2, with an average score of 2.45, characterizes users who encounter more significant challenges and exhibit incomplete satisfaction with their ChatGPT experience. These results demonstrate that while the majority of respondents possess moderate expectations for the use of ChatGPT, notable disparities exist among these clusters concerning expectations, SI, and available assistance, all of which impact their experience with this technology.

The analysis shown in Table 6 reveals some significant conclusions regarding the distinguishing characteristics of the two clusters identified in this study. Clusters 1 and 2 exhibit notable disparities regarding the motivation to resume utilizing ChatGPT. According to the rounding of the mean value of the cluster factors, Cluster 1 comprises a cohort of students with a heightened motivation to persist in utilizing ChatGPT. This is motivated by their favorable assessments of the technology's use, particularly for PE, EE, and FC.

Cluster 1 delineates a cohort of students who see that the use of ChatGPT offers distinct and readily attainable advantages underpinned by sufficient conditions. They possess elevated expectations regarding the efficacy of this technology, perceive ChatGPT as user-friendly, and believe there is adequate assistance during its utilization. This fosters their motivation to resume utilizing ChatGPT, signifying a favorable experience and contentment with the technology.

Conversely, Cluster 2 comprises students who need more motivation to resume utilizing ChatGPT. While they assessed the PE and EE dimensions as sufficiently satisfactory, they perceived that the remaining two elements, specifically SI and FC, did not facilitate optimal utilization of ChatGPT. Low scores in Social Influence suggest resistance to social encouragement for the usage of this technology. Still, low scores in FC reflect an absence of external features that facilitate the utilization of ChatGPT, such as infrastructure or accessibility. Cluster 2 was disinclined to utilize ChatGPT further due to insufficient support from their social environment and inadequate settings to optimize the utilization of this technology.

The disparity between these two groups underscores the significance of facilitating elements, including user-friendliness, explicit advantages, and sufficient social support and resources, in fostering interest and motivation to persist in utilizing technology like ChatGPT. Cluster 1, characterized by a favorable experience, showed a greater propensity to adopt and persist in using the technology. In contrast, Cluster 2, which perceived deficiencies in these areas, demonstrated diminished interest in future engagement with ChatGPT.

4. CONCLUSION

The clustering of objects according to specific criteria, as performed in this study to categorize Jambi University students based on their ChatGPT usage behavior, holds substantial significance and applicability in real-world contexts. This study's findings offer significant insights for universities, particularly on student interactions with ChatGPT technology and the determinants influencing their inclination to reuse it.

In this scenario, categorizing students into two distinct clusters – Cluster 1, comprising those motivated to reuse ChatGPT, and Cluster 2, consisting of those lacking such drive –

facilitates a more targeted development plan for each group. For students in Cluster 1, the institution can implement new features or offer further training to enhance their experience with ChatGPT, hence augmenting their satisfaction and the use of the technology. Conversely, for students in Cluster 2 who exhibit diminished interest in utilizing ChatGPT, universities may adopt an alternative strategy by identifying the obstacles that impede their engagement, encompassing both social factors and supportive conditions, and endeavoring to surmount these barriers to enhance their motivation for employing this technology.

Another practical implication is that these findings can be utilized by technology developers or firms offering AI-based services, such as ChatGPT, to enhance the user experience in a more personalized manner tailored to the specific needs of each user group. By comprehending the attributes and inclinations of each user cluster, developers may create more pertinent features and optimize user interactions more efficiently. These findings can serve as a foundation for developing policies that facilitate the integration of technology in education and enhance technology-based teaching methodologies to promote student adoption and optimal utilization of technology.

This study's results can inform policymakers in educational institutions, technology developers, and other stakeholders about user behavior patterns, enabling the optimization of technology applications across various contexts, thereby enhancing technology effectiveness and delivering more significant benefits to users.

5. ACKNOWLEDGMENTS

This research was effectively conducted with the help and facilitation of Jambi University, where the author participated as both a lecturer and a student.

REFERENCES

[1] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *J. Ind. Inf. Integr.*, vol. 23, no. May, p. 100224, 2021.

[2] OpenAI, "Introducing ChatGPT," *OpenAI*, 2023. [Online]. Available: https://openai.com/blog/chatgpt. [Accessed: 30-Nov-2022].

[3] B. Thormundsson, "ChatGPT - statistics & facts," *Statista*, 2023. [Online]. Available: https://www.statista.com/topics/10446/chatgpt/. [Accessed: 30-Nov-2022].

[4] Y. Shen *et al.*, "ChatGPT and Other Large Language Models Are Double-edged Swords," *Radiology*, vol. 307, no. 2, 2023.

[5] S. S. Gill *et al.*, "Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots," *Internet Things Cyber-Physical Syst.*, vol. 4, no. May 2023, pp. 19–23, 2024.

[6] Z. Y. Pamungkas and A. Sudiarno, "Implementasi Model UTAUT (Unified Theory of Acceptance and Use of Technology) untuk Menganalisis Faktor- Faktor yang Mempengaruhi Penggunaan Aplikasi Brimo," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 3, pp. 569–578, 2022.

[7] R. Piarna and F. Fathurohman, "Adopsi E-Commerce Oleh Konsumen Milenial Pada Produk Umkm E-Commerce Adoption By Millenial Consumers on Umkm Products in Subang City Using Utaut Model in Consumer Contex," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 1021–1028, 2020.

[8] Y. M. Saragih, E. Setiawan Panjaitan, R. Yunis, and P. Korespondensi, "Peran Interaktivitas Dalam Penggunaan E-Learning: Perluasan Model Utaut the Role of Interactivity in E-Learning Usage: an Extension of Utaut Model," vol. 10, no. 1, pp. 123–130, 2023.

[9] M. Nurhidayati and N. Khasanah, "Penggunaan Metode K-Means Cluster Untuk Mengklasifikasikan Kemampuan 4C Mahasiswa," *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 2, pp. 160–169, 2021.

[10] S. Rahmayani, S. Sumarno, and Z. A. Siregar, "Analysis of K-Means Algorithm for Clustering of Covid-19 Social Assistance Recipients," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 1, pp. 77–84, 2022.

[11] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019.

[12] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021.

[13] Fathurrahman, S. Harini, and R. Kusumawati, "Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (DBI)," *J. Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023.

[14] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviuddin, and A. Septiarini, "Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 2015–2027, 2023.

[15] A. Fira, C. Rozikin, and G. Garno, "Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Penyebaran Covid-19 di Indonesia," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 133–138, 2021.

[16] M. Wahyudi and L. Pujiastuti, "Komparasi K-Means Clustering dan K-Medoids dalam Mengelompokkan Produksi Susu Segar di Indonesia," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 243–254, 2022.

[17] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Q.*, vol. 27, no. 3, pp. 425–478, 2003.

[18] S. Ayem, E. K. Cahyaning, I. Ramadhan, M. Nurlitawati, H. Langkodi, and F. A. Trasno, "Faktor-Faktor Yang Mempengaruhi Behavior Intention Terhadap Penggunaan Digital Payment : Systematic Literature Review," *J. Pendidik. Ekon.*, vol. 12, no. 2, pp. 196–206, 2024.

[19] U. Merdekawati, D. M. K. Nugraheni, and O. D. Nurhayati, "Analisis Penerimaan dan Kesuksesan Aplikasi M-health pada Lansia menggunakan Model UTAUT dan Delone & McLean," *J. Sist. Inf. Bisnis*, vol. 14, no. 3, pp. 267–276, 2024.

[20] N. A. Ainul Bashir, "Penerapan Model UTAUT 2 Untuk Mengetahui Faktor-Faktor Yang Memengaruhi Penggunaan SIORTU," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 5, no. 1, pp. 42–51, 2020.

[21] A. H. Ramadhanty and Marwini, “Analisis Faktor-Faktor Penerimaan Dan Penggunaan Layanan Mobile Banking Syariah Sebagai Bukti Dari Model Utaut Yang Dimodifikasi,” *Diponegoro J. Islam. Econ. Bus.*, vol. 1, no. 3, pp. 185–194, 2021.

[22] F. N. S. Inggih Permana, “The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm,” *IJIRSE Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022.

[23] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri),” *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022.

[24] J. Mueller and L. Massaron, *Machine learning for dummies*. 2016.

[25] N. Fajriati and B. Prasetiyo, “Optimasi Algoritma Naive Bayes dengan Diskritisasi K-Means pada Diagnosis Penyakit Jantung,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 503–512, 2023.

[26] S. Monalisa, T. Nurainun, and M. Hartati, “Penerapan Algoritma K-Means Dan Metode Marketing Mix dalam Segmentasi Mahasiswa dan Strategi Pemasaran,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 61, 2021.

[27] A. P. Reynolds, G. Richards, B. De La Iglesia, and V. J. Rayward-Smith, *Clustering rules: A comparison of partitioning and hierarchical clustering algorithms*, vol. 5, no. 4. 2006.

[28] H. Lailatul Ramadhania, L. Zakaria, and D. Nusyirwan, “Aplikasi Metode Sillhouette Coefficient, Metode Elbow dan Metode Gap Staticstic dalam Menentukan K Optimal pada Analisis K-Medoids,” *J. Siger Mat.*, vol. 04, no. 01, pp. 1–10, 2023.

[29] S. Bahri and D. M. Midyanti, “Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 1, pp. 165–172, 2023.

[30] H. T. A. Simanjuntak, P. E. P. Silaban, J. K. S. Manurung, and V. H. Sormin, “Klasterisasi Berita Bahasa Indonesia Dengan Menggunakan K-Means Dan Word Embedding,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 641–652, 2023.

[31] F. A. Dewa and M. T. Jatipaningrum, “Segmentasi E-Commerce Dengan Cluster K-Means Dan Fuzzy C-Means (Studi Kasus: Media Sosial di Indonesia yang diunduh di Play Store),” *J. Stat. Ind. dan Komputasi*, vol. 4, no. 1, pp. 53–67, 2019.

[32] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987.