The variation of the data within a group is the indicator that shows the homogeneity of the group. Furthermore, the smallest standard deviation of the p variables for an object has the potential to be the medoid of a group. This is in line with the K-Means method, where the center is the average of the variables in the group. An object with a value close to or equal to the average will have a standard deviation of all variables which is smaller than others. Meanwhile, the smallest standard deviation for several objects can have the same value, though they originate from a data set with a high variation. This object set can become medoids, hence the FKM algorithm partitions objects and the data amount per object based on the standard deviation groups.

However, an empty group may occur when an identical object is selected as the medoid because the object is in the same group. This is overcome by adding a second step of object partitioning to the FKM algorithm which is carried out based on the sum of sorted variable values. The initial medoids are chosen from the representative object of the first $k$ blocks of a combination of standard deviation and sum of variable values. This process ensures that the selected initial medoids of identical objects will be represented by one of the objects. Although the purposes are not exactly the same, the concept of the sum data is also used in research about decision analysis [17].

**An illustrative example for selection of the initial medoids**

Suppose that 15 objects with two variables, X and Y, are grouped into three clusters. The initial medoids shown in Table 1 were obtained based on (3), (4), and step 1 of the Flexible K-Medoids. Furthermore, the SFKM method produced the initial medoids, namely objects I, J, and H. The non-medoid objects have the same distance to objects I and J because the two things are identical. Figure 1A showed that objects to be partitioned into three groups in the SFKM methods are classified into two. Meanwhile, the SKM method always uses the object I or J as one an initial medoids. Assuming the random result obtained are objects H, J, and I as one of the initial medoids, then these groups will have their own object members, as shown in Fig. 1B. The condition causes the SFKM and the SKM methods to be inflexible in choosing the candidate of the final medoids. The 15 objects are categorized into 12 blocks of standard deviation with the same sum of p variables, arranged from the smallest to the largest, if the proposed Flexible K-Medoids was used in the initial stage. The FKM method produces the initial medoids, namely objects A, C, and E, which create a composition of members group, including six, four, and five, respectively, hence there is flexibility to select an object as the final medoids as shown in Fig. 1C.

**Table 1.** Artificial Small Dataset (After Sorted by Standard Deviation and Sum)

| Object | X | Y | FKM (Proposed) | | SFKM | SKM |
|---|---|---|---|---|---|---|
| | | | Stdev | Sum | $(v_j)$ | $(a_j)$ |
| A | 47 | 46 | 0.7071 | **93*** | 1.1002 | 2685 |
| B | 46 | 47 | 0.7071 | 93 | 1.1002 | 2685 |
| C | 220 | 221 | 0.7071 | **441*** | 1.2046 | 2847 |
| D | 221 | 220 | 0.7071 | 441 | 1.2034 | 2843 |
| E | 180 | 175 | 3.5355 | **355*** | 0.8717 | 2197 |
| F | 15 | 21 | 4.2426 | 36 | 1.4164 | 3392 |
| G | 34 | 43 | 6.3639 | 77 | 1.1804 | 2859 |
| H | 120 | 103 | 12.0208 | 223 | **0.6777*** | 1859 |
| I | 140 | 122 | 12.7279 | 262 | **0.6583*** | **1818*** |
| J | 140 | 122 | 12.7279 | 262 | **0.6583*** | **1818*** |
| K | 80 | 104 | 16.9706 | 184 | 0.7801 | 2056 |
| L | 120 | 155 | 24.7487 | 275 | 0.7256 | 1937 |
| M | 150 | 100 | 35.3553 | 250 | 0.6926 | 1884 |
| N | 250 | 200 | 35.3553 | 450 | 1.2633 | 2980 |
| O | 271 | 219 | 36.7696 | 490 | 1.4671 | 3424 |

*Initial medoids

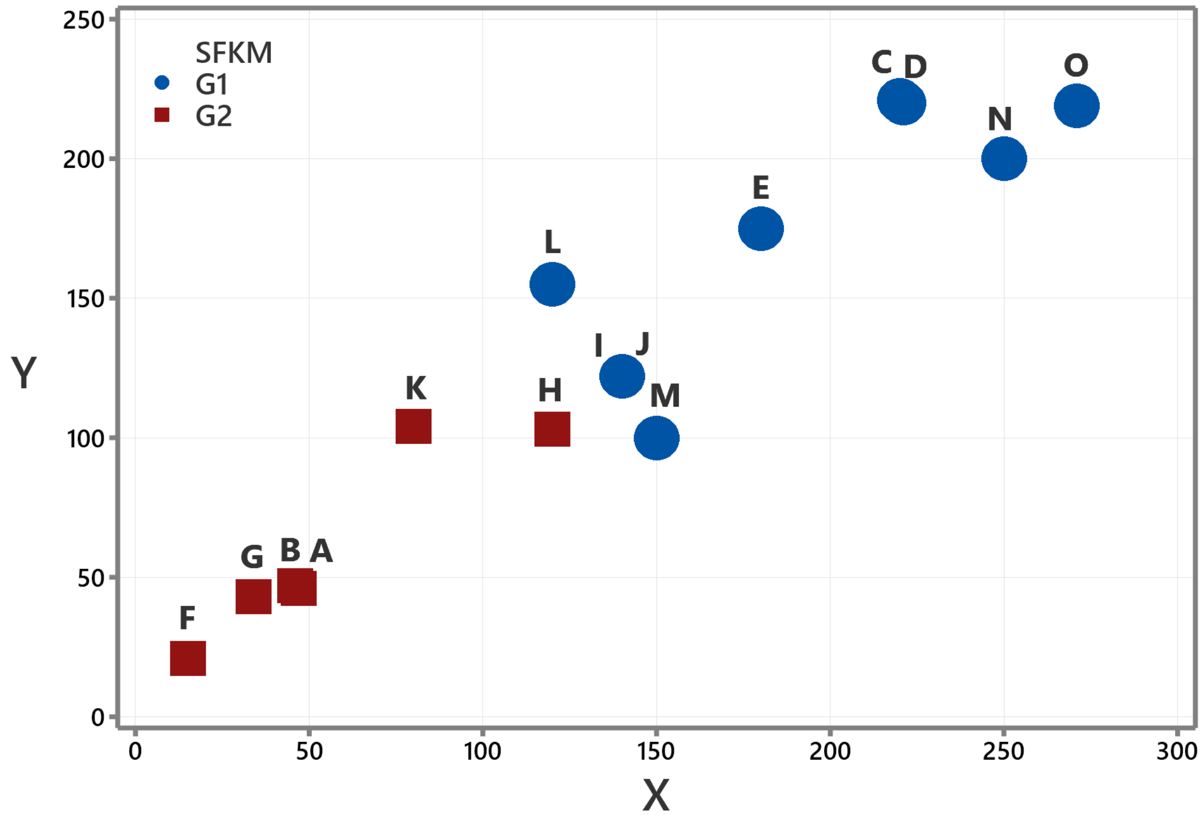


**Figure 1A.** Initial Groups by Simple and Fast K-Medoids (SFKM)

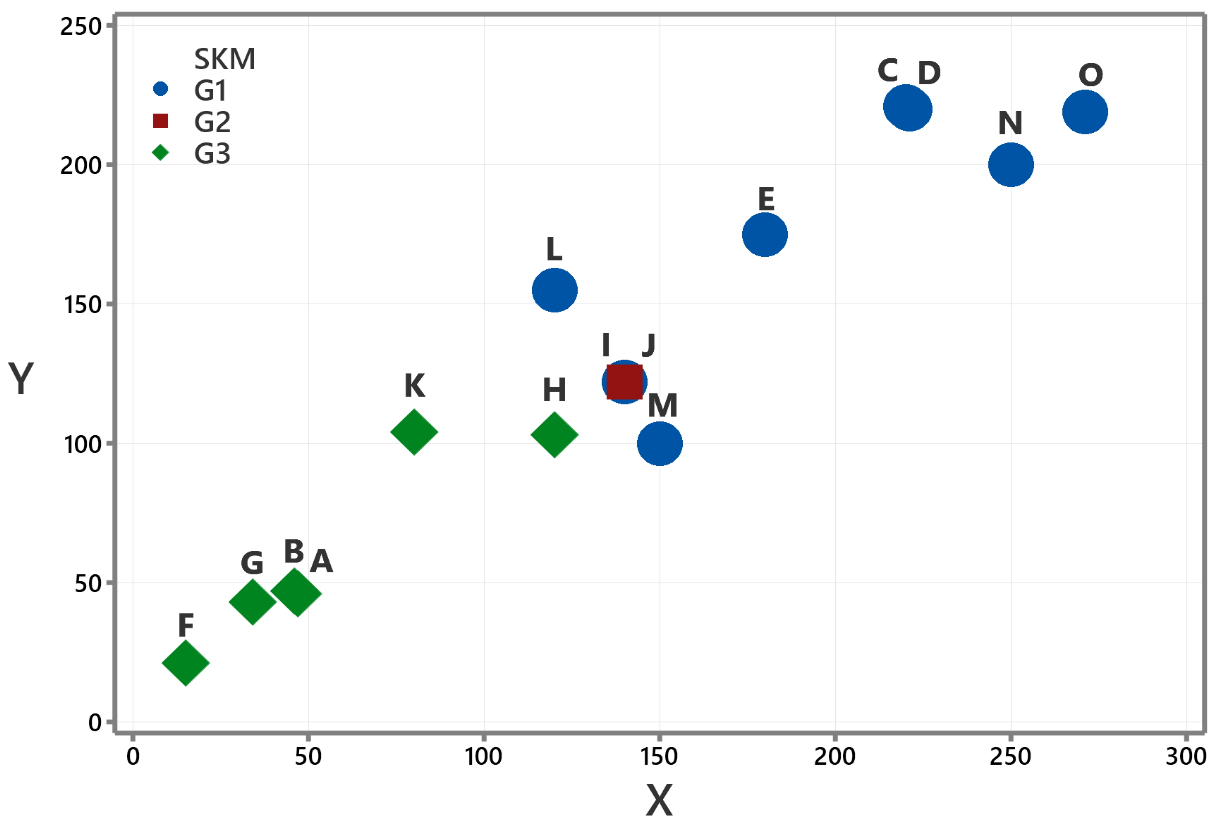


**Figure 1B.** Initial Groups by Simple K-Medoids (SKM)

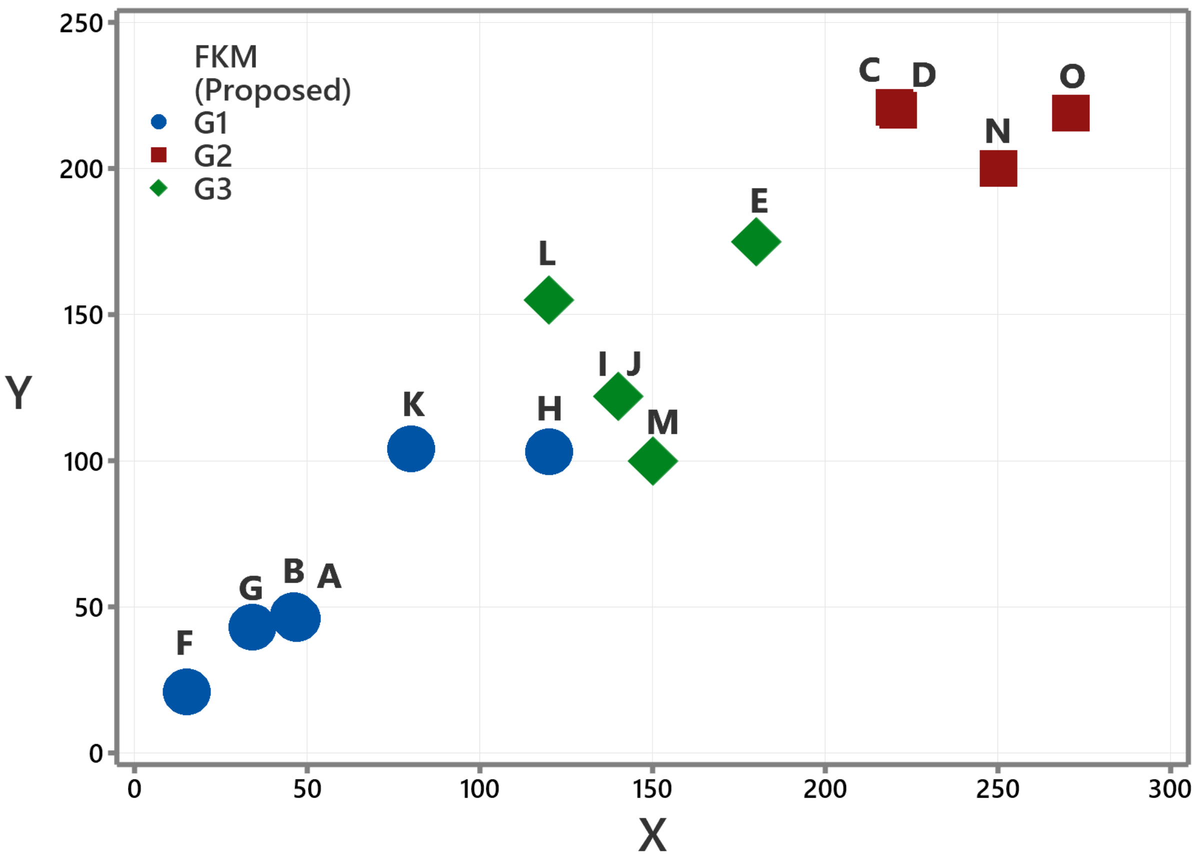


**Figure 1C.** Initial Groups by Flexible K-Medoids (proposed method)

### 2.2. Proximity Measure

The simple matching coefficient is used as a proximity measure to implement binary, categorical, and nominal variables [2, 18-19]. Furthermore, the differences between objects $i$ and $j$ for the variable $f$ of the simple matching without missing data are as follows,

$$d_{ij}^{f} = \begin{cases} 1 & if \;\; x_{if} \neq x_{jf} \\ 0 & if \;\; x_{if} = x_{jf} \end{cases} \tag{6}$$

The Euclidean distance is used for numerical data, where the distance of the objects $i$ and $j$ are defined as follows,

$$d(i,j) = \left[\sum_{l=1}^{p} \left(x_{il} - x_{jl}\right)^2\right]^{1/2} \tag{7}$$

This research applied the Generalized Distance Function (GDF) between objects $i$ and $j$ for non-missing mixed data as follows [2, 20],

$$d(i,j) = \sum_{s=1}^{p_b} \delta_b\left(x_{is}, x_{js}\right) + \sum_{t=1}^{p_c} \delta_c\left(x_{it}, x_{jt}\right) + \sum_{r=1}^{p_n} \delta_n\left(x_{ir}, x_{jr}\right). \tag{8}$$

where $p_b$, $p_c$, and $p_n$ are the number of binary, categorical, and numeric (interval and ratio) variables, respectively. However, $\delta_b\left(x_{is}, x_{js}\right)$ and $\delta_c\left(x_{it}, x_{jt}\right)$ are the simple matching distance, while $\delta_n\left(x_{ir}, x_{jr}\right)$ is the Euclidean distance.

The standardized method is proposed to avoid an influence from the attribute values dimension using Eq. (9) or (10) below,

$$Z_{li} = f.\left(\frac{r_{li} - r_{l1}}{r_{lm} - r_{l1}}\right); \;\; i = 1, 2, \ldots, n \tag{9}$$

where $r_{li}$ is the rank of object $ith$ variable $l$, and $f$ is the transformation multiplier for standardization. Rank-based transformation is widely used in non-parametric statistics [21]. To reduce the chance of losing information due to decreasing the numerical to the ordinal scale, for numerical data, we reformulate as follows,

$$Z_{li} = f.\left(\frac{X_{li} - min(x_l)}{max(x_l) - min(x_l)}\right); \;\; i = 1, 2, \ldots, n. \tag{10}$$

Equation (9) or (10) is used in some cases of mixed data between nominal and ordinal or between numeric and categorical data, before calculating certain distances.

### 2.3. Artificial and Real Datasets

#### 2.3.1. Artificial Data

The performance of the proposed method is evaluated by generating some artificial data sets. Furthermore, five experiments are constructed, having a cluster size of two, four, five, six, and nine with categorical, numerical, and mixed data. This was repeated fifty times for each experiment.

2.3.2. Real Data

The effectiveness of this method is determined by Irvine (UCI) repository [22], including iris, ionosphere, soybean small, primary tumor, Heart Disease (HD) case 1, and zoo data. Table 2 shows the profile of the six real datasets and a brief description was presented in sub-section 3.2.

**Table 2.** Profile of the real dataset

| Data Set | $n$ | $p_n$ | $p_c$ | $k$ | Type |
|---|---|---|---|---|---|
| Iris | 150 | 4 | - | 3 | Numerical |
| Ionosphere | 351 | 33 | - | 2 | Numerical |
| Soybean small | 47 | - | 35 | 4 | Categorical |
| Primary tumor | 65 | - | 17 | 4 | Categorical |
| HD case 1 | 294 | 5 | 8 | 5 | Mixed |
| Zoo | 101 | 1 | 15 | 7 | Mixed |

$n$: number of objects; $p_n$: number of numerical variables; $p_c$: number of categorical variables; $k$: number of actual clusters

### 2.4. Evaluation Indexes

The Adjusted Rand Index (ARI) was used to compare the performance of the proposed method. Suppose that U is the true partition and V is a clustering result under consideration, then [23] formulated ARI as follows:

$$ARI = \frac{2(ad-bc)}{(a+b)(b+d)+(a+c)(c+d)} \quad (11)$$

where $a$ is the number of pairs of the objects that are placed in the same group in U and in the same class in V; $b$ is the number of pairs in the same cluster in U but not in the same group in V, $c$ is the number of pairs in the same cluster in V but not in the same class in U, and $d$ is the number of pairs in different groups in U and different classes in V. The clustering algorithm efficiency was evaluated using indexes accuracy defined as

$$AC = \frac{1}{n}\sum_{g=1}^{k} a_g \quad (12)$$

where $n$, $k$ and $a_g$ are the number of objects, clusters, and objects from the considered groups that are correctly assigned to the true clusters, respectively.

## 3. Results and Discussion

### 3.1. Experiment Results of Artificial Datasets

Figure 2 illustrates an example of the artificial data set generated for four groups (Fig. 2A) and nine groups (Fig. 2B). Table 3 shows the performance of the initial and final groups based on adjusted Rand index. Although it is uncommon to compare ARI based on initial groups, the performance on various data types showed good results, and no empty initial groups were discovered. Figure 3 demonstrated the minimum, average, and maximum values of the ARI from the fifty times trial for the final groups of each experiment. The average ARI of categorical data types with a group size of two was 98% with a standard deviation of 0.025.

Meanwhile, mixed data with a group size of six had the lowest ARI value of 78% with a standard deviation of 0.152. The overall performance of the initial medoids for various data types was good and guarantees that the group is not empty. This group makes it easy to find a medoid to produce a final group with high clustering accuracy.

### 3.2. Real Datasets

The proposed method, [7] and [8] was used to determine the initial medoids in this section. Furthermore, one sample of random results is used for the SKM method, assuming the result obtained is the $k$ first smallest at the value of $a_j$ from (2). This constraint was applied for all real datasets and different random results that may be better are obtained for the SKM method.

**Table 3.** Average and deviation standard of adjusted Rand index based on initial and final groups for artificial dataset

| No. | Type | $n$ | $k$ | $p$ | $p_b$ | $p_o$ | $p_n$ | Average of ARI for initial groups | Std deviation of ARI for initial groups | Average of ARI for final groups | Std deviation of ARI for final groups |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Categorical | 157 | 2 | 16 | 16 | - | - | 0.754 | 0.037 | 0.984 | 0.025 |
| 2. | Numerical | 340 | 4 | 2 | - | - | 2 | 0.426 | 0.040 | 0.894 | 0.127 |
| 3. | Categorical | 500 | 5 | 6 | 6 | - | - | 0.649 | 0.006 | 0.989 | 0.050 |
| 4. | Mixed | 608 | 6 | 12 | 3 | 1 | 8 | 0.053 | 0.005 | 0.781 | 0.152 |
| 5. | Numerical | 657 | 9 | 2 | - | - | 2 | 0.346 | 0.033 | 0.875 | 0.037 |

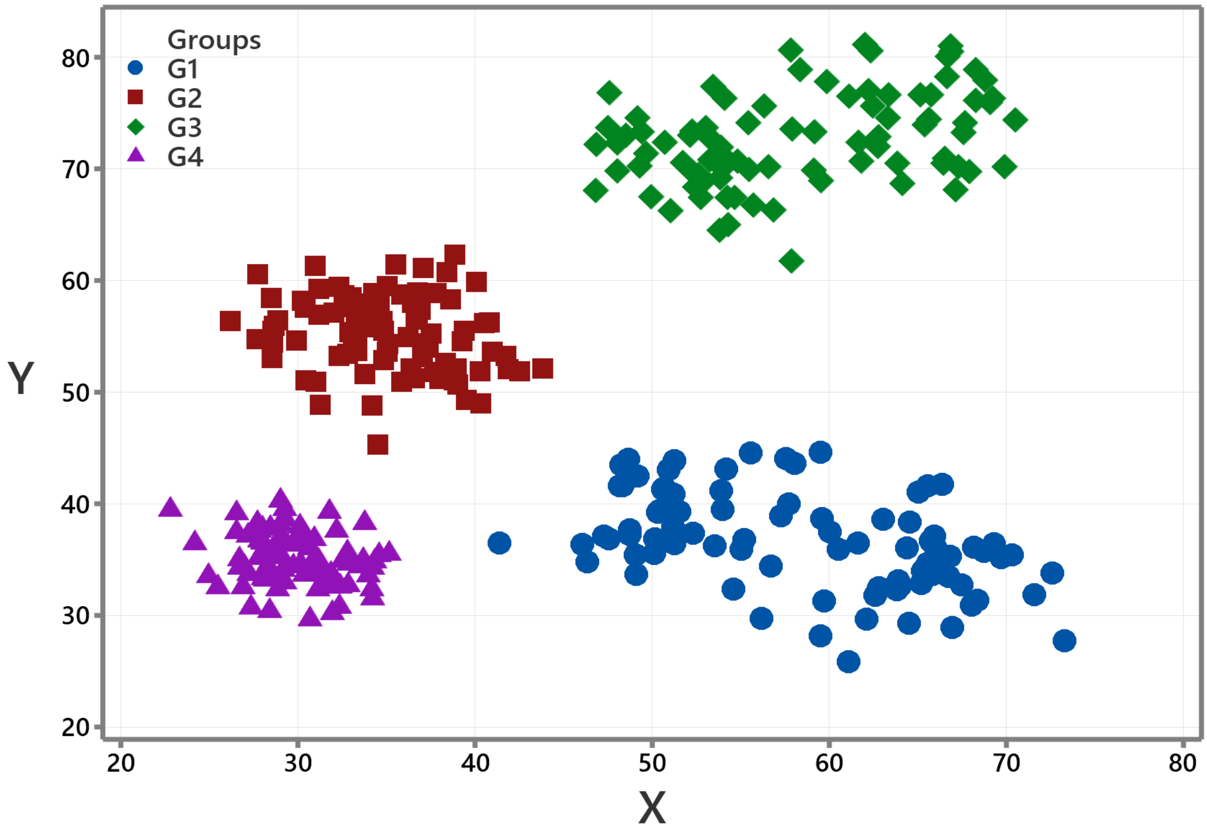


**Figure 2A.** Example of the artificial data set for four groups

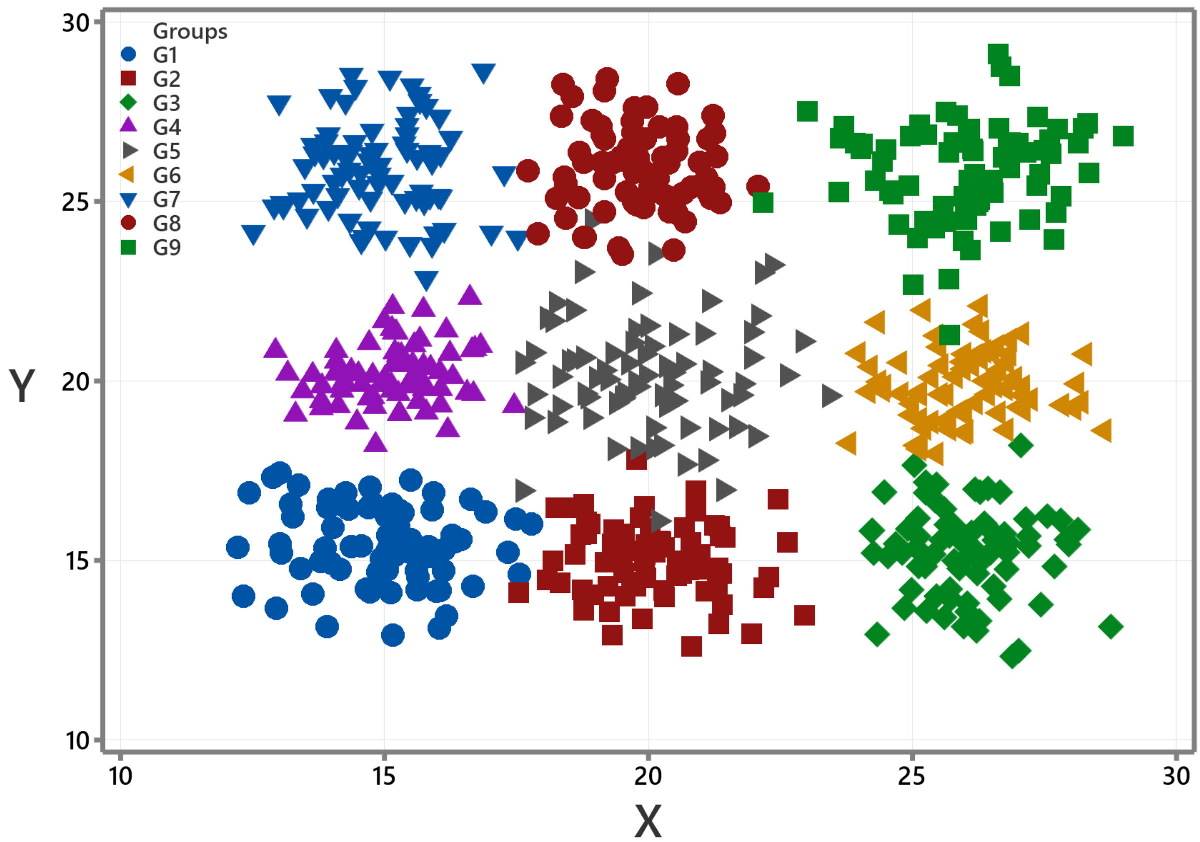


**Figure 2B.** Example of the artificial data set for nine groups

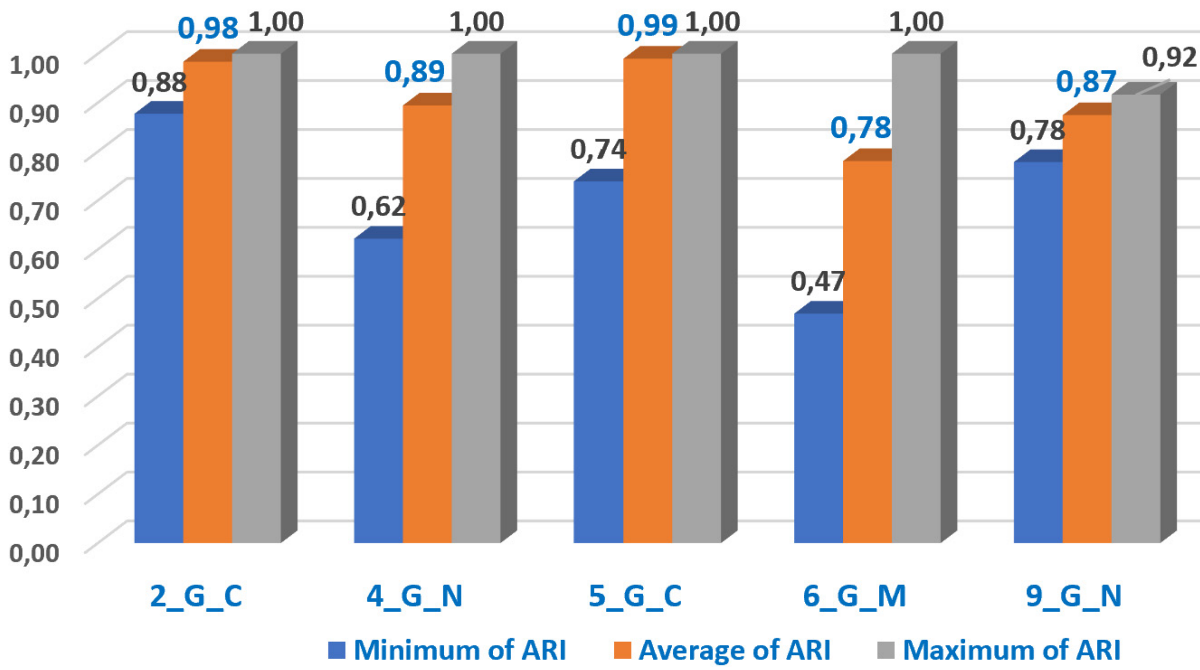


**Figure 3.** Profile of adjusted Rand index for artificial data set at final groups

3.2.1. Iris Data

The data contain 150 objects assigned to three classes with four numerical variables. The SFKM, SKM, and proposed method on the iris data produce an initial groups that classifies data into three clusters. The final medoids is determined by the members of each initial group. FKM method produces a clustering accuracy based on the initial medoids better than SFKM and SKM. The accuracy based on the initial medoids of the new method is equal to the accuracy based on the final medoids in the SKM method. In the iris data, the composition of the initial medoid members from the random results of the SKM method is different from the proposed method. This is because the initial medoid members of the proposed method lacks objects with the smallest distance from others. However, the object must be included as an initial medoids in the SKM method. Table 4 depicted the number of group members from the three methods.

**Table 4.** The number of initial groups members and accuracy for iris data

| Cluster and Accuracy | SFKM | SKM | FKM (proposed) |
|---|---|---|---|
| 1. Cluster 1 | 72 | 71 | 54 |
| 2. Cluster 2 | 2 | 9 | 43 |
| 3. Cluster 3 | 76 | 70 | 53 |
| 4. Accuracy (%) | 67.33 | 70.00 | **95.3** |
| 5. Adjusted RI | 0.426 | 0.445 | **0.866** |

3.2.2. Ionosphere Data

The ionosphere data contains 351 instances with 34 numerical variables assigned to two clusters. In this research, the second variable is not included because the data has the same value across all objects. Table 5 shows the initial groups profile based on the initial medoids for three methods using the limitation of one random result, namely the two smallest values of results (2) for the SKM method. The initial medoid members of the FKM method does not contain the most central object, as with iris data. Therefore, the results of the SKM method's initial medoids arrangement may not be the same as the proposed method. When compared with the SFKM method, one initial medoid sample has the same members. In addition, the quality of accuracy and ARI of the proposed method is better than that of the SFKM and SKM methods.

**Table 5.** The number of initial groups members and accuracy for ionosphere data

| Cluster and Accuracy | SFKM | SKM | FKM (proposed) |
|---|---|---|---|
| 1. Cluster 1 | 202 | 202 | 50 |
| 2. Cluster 2 | 149 | 149 | 301 |
| 3. Accuracy (%) | 63.53 | 63.53 | **78.34** |
| 4. Adjusted RI | 0.069 | 0.069 | **0.291** |

3.2.3. Soybean Small Data

The adjusted Rand index of the initial groups from the FKM outperforms the SFKM and SKM methods in the soybean small data, by positioning 35 variables as categorical data. These three methods had a similar initial medoids member object. However, the FKM method does not contain the most central object as the initial medoids. The quality of the initial group produced by the three methods in the soybean small data was adequate for the final medoids selection. Table 6 shows the number of initial group members for soybean small data.

**Table 6.** The number of initial groups members and accuracy for soybean small data

| Cluster and Accuracy | | SFKM | SKM | FKM (proposed) |
|---|---|---|---|---|
| 1. | Cluster 1 | 7 | 7 | 8 |
| 2. | Cluster 2 | 6 | 6 | 14 |
| 3. | Cluster 3 | 16 | 16 | 5 |
| 4. | Cluster 4 | 18 | 18 | 20 |
| 5. | Accuracy (%) | 36.2 | 36.2 | **51.1** |
| 6. | Adjusted RI | 0.098 | 0.098 | **0.527** |

3.2.4. Primary Tumor Data

There are 339 cases of primary tumor data including missing data with 17 categorical ones and are classified into 22 classes. This research used 65 non-missing objects from four groups, namely groups 1, 2, 4, and 5. Several objects had the same value on all variables, hence are identical. The application of the SFKM algorithm produces two empty initial groups, which are caused by the second, third and fourth medoids being identical objects. However, the SFKM method does not explain the solution to this problem. This object arrangement may also be a random result of the SKM method. The FKM method ensures that identical objects will be in one group, though the thing is either the initial or final medoids. Table 7 listed the initial group members of the three methods. The adjusted Rand index of the proposed method based on the initial medoids is 0.2098 with a clustering accuracy is 55.38%. Meanwhile, a sample from the SKM method produces a clustering accuracy of 50.77% with an adjusted Rand index of 0.168. This initial medoid members of SKM and SFKM methods are the same.

**Table 7.** The detailed number of initial group members for primary tumor data

| True Cluster | SFKM algorithm | | | | SKM algorithm*) | | | | FKM (proposed) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
| Cluster 1 | 13 | 15 | - | - | 13 | 15 | 0 | 0 | **20** | 5 | 1 | 2 |
| Cluster 2 | 0 | 12 | - | - | 0 | 12 | 0 | 0 | 2 | **5** | 0 | 5 |
| Cluster 3 | 7 | 1 | - | - | 7 | 0 | 1 | 0 | 3 | 1 | **4** | 0 |
| Cluster 4 | 16 | 1 | - | - | 16 | 0 | 0 | 1 | 3 | 3 | 4 | **7** |
| $n_g$ | 36 | 29 | 0 (empty) | 0 (empty) | 36 | 27 | 1 | 1 | 28 | 14 | 9 | **14** |

$n_g$: the number of groups members, *) example of probably the random results for initial medoids

**Table 8.** The number of initial groups members and accuracy for Heart Disease Case 1 data

| Cluster and Accuracy | SFKM | SKM | FKM (proposed) |
|---|---|---|---|
| Cluster 1 | 57 | 68 | 169 |
| Cluster 2 | 71 | 53 | 19 |
| Cluster 3 | 44 | 58 | 13 |
| Cluster 4 | 77 | 53 | 31 |
| Cluster 5 | 45 | 62 | 62 |
| Accuracy (%) | 27.21 | 31.97 | **57.14** |
| Adjusted RI | 0.010 | 0.054 | **0.318** |

3.2.5. Heart Disease Case 1 Data

The heart disease case 1 data contained 294 instances with eight categorical and five numerical data, assigned to 5 groups. Equation (9) was applied for numerical data with $f = 5$ before being executed with (7) and (6) for categorical data, then it was used to create (8). The initial medoids profile was obtained, such as Table 8, and the proposed method produced an adjusted Rand index and clustering accuracy that is better than SFKM and one sample of SKM.

3.2.6. Zoo Data

There are several identical objects in a zoo dataset which consists of 101 objects. The initial medoids produced by the SFKM method contains two pairs of things, which are similar. The first pair ranks fourth and fifth smallest, while the second pair ranks sixth and seventh. This causes the two initial groups to be empty, as shown in Table 9, making the next stage non-feasible. The initial medoids random results from the SKM method can produce a combination of the object that are identical, as in the SFKM method. Moreover, the zoo data have the same distance spread over 19 blocks. There is a block containing 10 objects with the same distance. The two blocks contain six objects each and five objects equidistant from the others. Four objects have the same distance as the two blocks, and the four blocks contain three evenly spaced objects each. Also, ten pairs of objects were observed with the same distance to others. These blocks produced by GDF with simple matching for binary variables and square euclidean distance for one numerical data were standardized using (10) with $f = 1$.

Table 10 demonstrates that random results may be obtained as the initial group of the SKM method using the same constraint. There are five initial medoid members of SKM which are the same as the SFKM, and two of them are identical objects, hence causing unbalanced initial group members. The zoo data has several identical data blocks, hence the iterative process to determine six initial medoids in the SKM method can lead to results containing identical objects. As with the primary tumor data, the proposed method can handle well zoo data. Table 11 showed the initial medoids selection from the Flexible K-Medoids method resulted in group members' composition.

The proposed method appropriately approaches data from two real cases, zoo and primary tumor data, and ensures that identical data are in one group. Furthermore, the flexible k-medoids outperforms other methods in small or large data that may contain data blocks with identical distances. Although the adjusted Rand index and clustering accuracy are the primary goal of the whole algorithm, the ARI from the initial medoids of the proposed method outperforms others, such as Fig. 4. The adjusted Rand index for SFKM and SKM methods for primary tumor and zoo data are not presented in Fig. 4 due to the empty initial groups produced by the initial medoids, though this value can be calculated.

**Table 9.** The detailed number of initial group members for zoo data based on the SFKM Algorithm

| **True Cluster** | **Simple and Fast K-Medoids Algorithm** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** | **Cluster 6** | **Cluster 7** |
| Cluster 1 | **32** | 0 | 0 | 0 | - | 9 | - |
| Cluster 2 | 0 | **14** | 0 | 6 | - | 0 | - |
| Cluster 3 | 0 | 2 | **2** | 1 | - | 0 | - |
| Cluster 4 | 0 | 0 | 0 | **13** | - | 0 | - |
| Cluster 5 | 0 | 0 | 0 | 4 | - | 0 | - |
| Cluster 6 | 0 | 0 | 8 | 0 | - | 0 | - |
| Cluster 7 | 0 | 3 | 1 | 6 | - | 0 | - |
| **The number of groups members** | 32 | 19 | 11 | 30 | 0 (empty) | 9 | 0 (empty) |

**Table 10.** The detailed number of initial group members for zoo data based on the SKM Algorithm

| True Cluster | Simple K-Medoids Algorithm | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** | **Cluster 6** | **Cluster 7** |
| Cluster 1 | **21** | 1 | 19 | 0 | 0 | 0 | 0 |
| Cluster 2 | 0 | 0 | 0 | **14** | 3 | 3 | 0 |
| Cluster 3 | 0 | 0 | 0 | 1 | **2** | 1 | 1 |
| Cluster 4 | 0 | 0 | 0 | 0 | 0 | **13** | 0 |
| Cluster 5 | 0 | 0 | 0 | 0 | 0 | 4 | 0 |
| Cluster 6 | 0 | 0 | 0 | 7 | 0 | 0 | 1 |
| Cluster 7 | 0 | 0 | 0 | 3 | 1 | 5 | **1** |
| **The number of groups members** | 21 | 1 | 19 | 25 | 6 | 26 | 3 |

**Table 11.** The detailed number of initial group members for zoo data based on the FKM Algorithm (Proposed)

| True Cluster | Flexible K-Medoids Algorithm (Proposed) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** | **Cluster 6** | **Cluster 7** |
| Cluster 1 | **40** | 0 | 1 | 0 | 0 | 0 | 0 |
| Cluster 2 | 0 | **16** | 1 | 3 | 0 | 0 | 0 |
| Cluster 3 | 0 | 1 | **0** | 4 | 0 | 0 | 0 |
| Cluster 4 | 0 | 0 | 4 | **9** | 0 | 0 | 0 |
| Cluster 5 | 0 | 0 | 1 | 3 | **0** | 0 | 0 |
| Cluster 6 | 0 | 0 | 0 | 0 | 0 | **8** | 0 |
| Cluster 7 | 0 | 0 | 0 | 0 | 3 | 3 | **4** |
| **The number of groups members** | 40 | 17 | 7 | 19 | 3 | 11 | 4 |

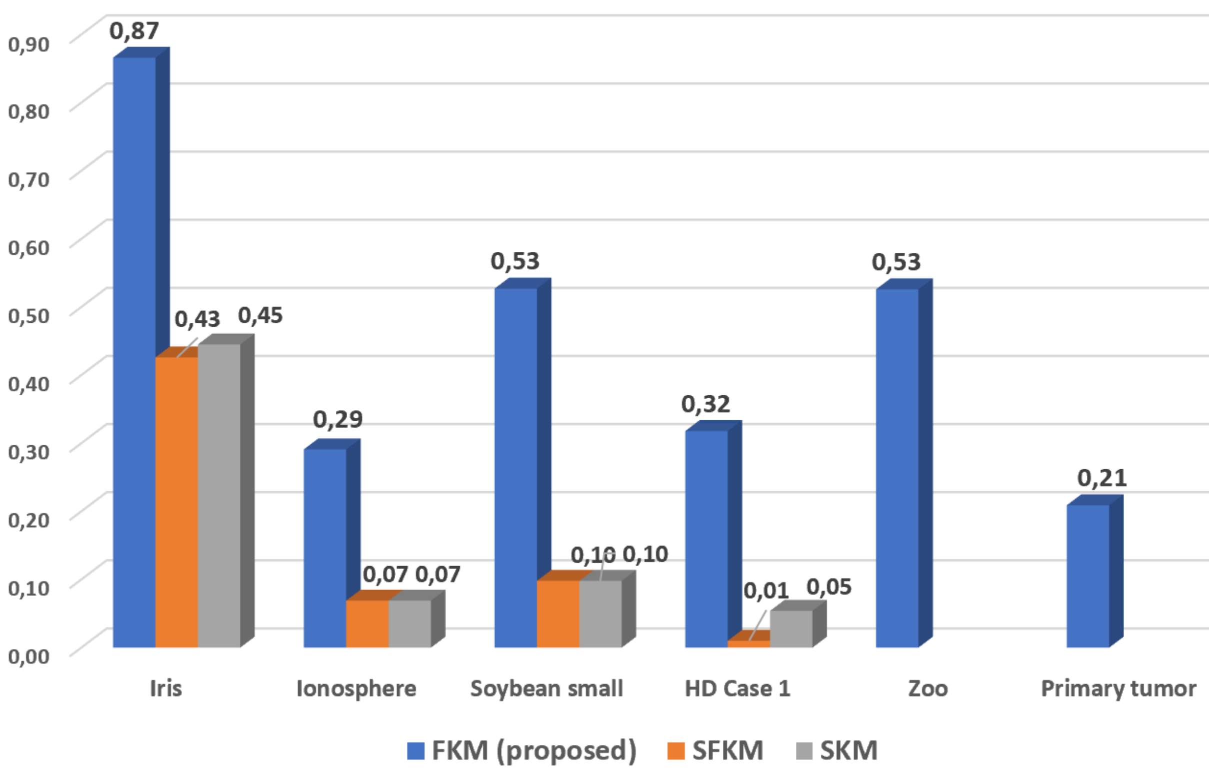


**Figure 4.** The adjusted Rand index of initial groups

3.2.7. Evaluation of Final Groups for Real Datasets

Figure 5 illustrates the comparison between the adjusted Rand index of the new method and others [7,8,12, 24-26]. Although it is unfair to compare with other methods because the distances used may vary, from the graph of the ARI the proposed method outperforms others. The adjusted Rand index for primary tumor data was not presented because there is an empty group produced by the SFKM method.

Table 12 depicts the detail of adjusted Rand index values and clustering accuracy from the proposed methods and several others [7,8,12,25,26]. The SFKM method was applied for ionosphere and heart disease case 1, with the same data used to obtain ARI and clustering accuracy. Overall, the performance of the adjusted Rand index and clustering accuracy of the proposed method for six real datasets is better than others.

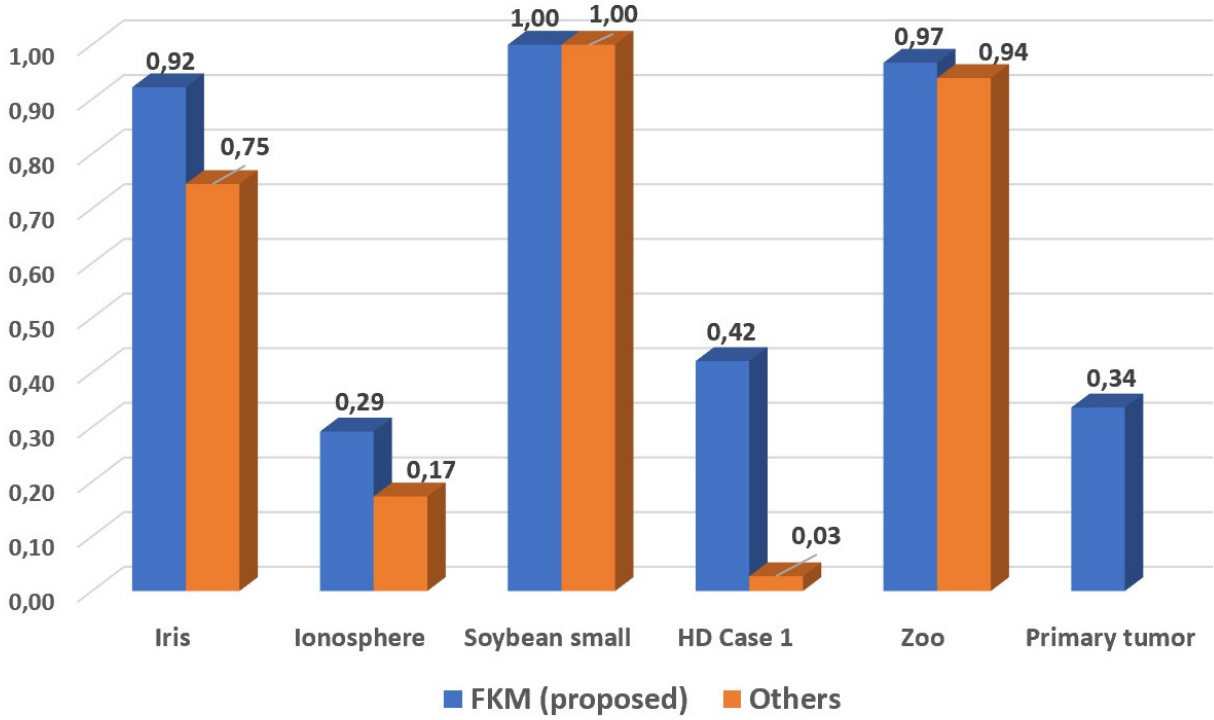


**Figure 5**. The adjusted Rand index of final groups

**Table 12.** Comparison adjusted Rand index and clustering accuracy of final groups for real datasets between FKM (proposed) and other methods

| Real Datasets | Adjusted Rand index | | Clustering accuracy (in percent) | |
|---|---|---|---|---|
| | FKM | Other methods | FKM | Other methods |
| 1. Iris | **0.922** | 0.745[12] | **97.3** | 92.0[7]; 95.3[8] |
| 2. Ionosphere | **0.292** | 0.173[12] | **78.35** | 69.2[7] |
| 3. Soybean small | **1.000** | 1.000[8,12] | **100** | 80.9[7]; 100[8]; 97.8[24]; 98.9[25] |
| 4. Primary tumor | **0.336** | - | **68.18** | - |
| 5. Heart disease case 1 | **0.421** | 0.027[7] | **63.61** | 26.2[7] |
| 6. Zoo | **0.967** | 0.901[25]; 0.939[26] | **96.04** | 82.2[8]; 88.8[25]; 89.9[26] |

# 4. Conclusion

The k-medoids algorithm has been extensively utilized and improved for data partitioning. This research proposes a method to obtain the initial medoids based on the representative object from a block of combination deviation and the sum of p variables values. Similar or identical objects are in a block. The principle of this method is that the relative position of an object alters with a different amount even though the deviation is the same. This new method of Flexible K-Medoids is applicable to all data types and it ensures that no initial groups is empty. The adjusted Rand index and clustering accuracy of experiments using artificial and six real data sets from the UCI repository show the good performance of FKM. This method provides another view of obtaining the initial medoids.

# REFERENCES


[1] Leonard K., P.J. Rousseeuw, "Clustering by Means of Medoids," 1987. https://www.researchgate.net/publication/ 243777819_Clustering_by_Means_of_Medoids

[2] Leonard K., P.J. Rousseeuw, "Finding groups in data: An introduction to cluster analysis," John Wiley & Sons, Inc., New York, 1990, pp. 28-37.

[3] Xindong W., Vipin. K., J. Ross Q., Joydeep G., Qiang Y., Hirosha M., Geoffrey J.M., Angus Ng, Bing L., Philip S.Y., Zhi-Hua Z., Michael S., David J.H., Dan S., "Top 10 algorithms in data mining," in Knowledge and Information Systems, 14, 1-37, 2008. DOI: 10.1007/s10115-007-0114-2

[4] Douglas S., M.J. Brusco, "Initializing K-means Batch Clustering: A Critical Evaluation of Several Techniques," in Journal of Classification, 24, 99-121, 2007. DOI: 10.1007/s00357-007-0003-0

[5] Malay K.P., "A Modified k-means Algorithm to Avoid Empty Clusters," in International Journal Recent Trends Engineering, 1, 221-226, 2009. https://www.researchgate.net/publication/ 228414762_A_ Modified_kmeans_Algorithm_to_Avoid_Empty_Clusters

[6] Zhexue H., "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values," in Proceedings of the First Pacific-Asia Conference on Knowledge Discoveray and Data Mining, Sprenger: Berlin, Germany, 21-34, 1997. https://citeseerx.ist.psu.edu/viewdo c/ summary?doi=10.1.1.94.9984

[7] Hae-Sang P., C. Jun, "A Simple and Fast Algorithm for K-Medoids Clustering," in Expert System with Applications, 36 (2), 3336–3341, 2009. DOI: 10.1016/j.eswa.2008.01.039

[8] Weksi B., F. Leisch, "Simple k-medoids partitioning algorithm for mixed variable data," in Algorithms, 12, 177, 1-15, 2019. DOI: 10.3390/a12090177

[9] Seyed M.R.Z, Mehdi M, F. Sadoughi, "Ranked k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets," in Knowledge-Based System, 39, 133-143, 2013. DOI: 10.1016/j.knosys.2012.10.012

[10] Erich S., P.J. Rousseeuw, "Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms," in Information Systems, Elsevier, 101, 101804, pp. 1-19, 2021. DOI: 10.1016/j.is.2021.101804.

[11] Anton V.U., I. Vasilyev, "Near-optimal large-scale k-medoids clustering," in Information Sciences, 545, 344-362, 2021. DOI: 10.1016/j.ins.2020.08.121

[12] Donghua Y., Guojun L., Mauzo G., X. Liu, "An improved k-medoids algorithm based on step increasing and optimizing medoids," in Expert System With Applications, 92, 464-473, 2018. DOI: 10.1016/j.eswa.2017.09.052.

[13] Shaker K.A., Z.N. Azeez, "Modify k-medoids algorithm with new efficiency method for biometric database classification," in International Journal of Computer Science and Mobile Computing, Vol. 4, Issu.10, 364-376, 2015,https://www.researchgate.net/publication/329922298 _Modify_k-medoids_Algorithm_with_New_Efficiency_M ethod_ for_Biometric_Database_Classification

[14] V. Thambusany, A. Dharmarajan, "Efficiency of k-means and k-medoids clustering algorithms using lung cancer dataset," in International Journal of Data Mining Techniques and Applications, vol. 05, issue 02, pp. 150-156, ISSN: 2278-2419, 2016. DOI: 10.20894/IJDMTA.102.005 .002.011

[15] Wijayaningrum V.N., P.N. Nur, "An improved crow search algorithm for data clustering," in EMITTER International Journal Engineering Technology, 8(1), 2020. DOI: 10.24003/emitter.v8i1.498.

[16] Nitesh S., Bharat C., A. Vasant, "An improved K-medoids clustering approach based on the crow search algorithm," in Journal of Computational Mathematics and Data Science, 3, 100034, 2022. DOI: 10.016/j.jcmds.2022.100034.

[17] Shahryar S., Yasaman P., "Modified Weighted Sum Method for Decisions with Altered Sources of Information," in Mathematics and Statistics, 7(3), 57-60, 2019. DOI: 10.13189/ms.2019.070301.

[18] Brian S.E., Sabine L., Morven L., D. Stahl, "Cluster analysis," 5th edn., John Wiley & Sons., Ltd., Publication, 2011, pp. 43-68.

[19] Rui X., D.C. Wunsch II, "Clustering," John Wiley & Sons, Inc., Hoboken, New Jersey, 2009, pp. 15-29.

[20] J.C. Gower, "A general coefficient of similarity and some of its properties," in Biometrics, vol. 27, no. 4, 857-871, 1971. DOI: 10.2307/2528823

[21] Adejumo T. Joel., Omonijo D. Ojo, Owolabi A. Timothy, Okegbade A. Ibukun, Odukoya A. Jonathan, Ayedun C. Ayedun, "Comparison of Rank Transformation Test Statistics with Its Nonparametric Counterpart Using Real-Life Data," Mathematics and Statistics, Vol. 8, No. 6, pp. 683 - 692, 2020. DOI: 10.13189/ms.2020.080608.

[22] M. Lichman, UCI Machine Learning Repository, University of California: Irvine, CA, USA, 2013. Avaliable online http://archive.ics.uci.edu/ml (accessed Nov. 10, 2021 and June. 15, 2022)

[23] L. Hubert, P. Arabie, "Comparing Partition," in Journal of

Classification, 2, 193-218, 1985. DOI: 10.1007/BF01908075

[24] F. Yuan, Y. Yang, T. Yuan, "A dissimilarity measure for mixed nominal and ordinal attribute data in k-Modes algorithm," in Applied Intelligence, 50, 1498-1509, 2020. DOI: 10.1007/s10489-019-01583-5

[25] J. Ji., W. Pang, Z. Li, F. He, G. Feng, X. Zhao, "Clustering mixed numeric and categorical data with cuckoo search," in IEEE Access, vol. 8, pp 30988-31003, 2020. DOI: 10.1109/ACCESS.2020.2973216

[26] J. Ji, R. Li, W. Pang, F. He, G. Feng, X. Zhao, "A Multi-View Clustering Algorithm for Mixed Numeric and Categorical Data," in IEEE Access, vol. 10, pp 24913-24924, 2021. DOI: 10.1109/ACCESS.2021.3057113