African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 42-54)
www.abjournals.org

NEW K-MEANS CLUSTERING METHODS THAT MINIMIZES THE TOTAL INTRA-CLUSTER VARIANCE

Eric U. Oti[1*], Sidney I. Onyeagu[2], Chike H. Nwankwo[2], Waribi K. Alvan[3] and George A. Osuji[2]

[1]Department of Statistics, Federal Polytechnic, Ekowe Bayelsa State, Nigeria
[2]Department of Statistics, Nnamdi Azikiwe University, Awka Anambra State, Nigeria
[3]Department of Physics with Electronics, Federal Polytechnic, Ekowe Bayelsa State, Nigeria
[*]Corresponding Author: eluchcollections@gmail.com (+2348037979262)

ABSTRACT: *In this paper, we present new k-means clustering methods namely: the modified k-means method and the enhanced k-means method. The modified k-means clustering method proposed updates cluster centroids depending on if a point is added to a cluster or a point is removed from a cluster; while the enhanced k-means clustering method uses the Minkowski's distance as its metric in a normed vector space instead of the usual Euclidean distance used in the modified k-means method and the existing methods. K-means clustering is one of the simplest and popular unsupervised learning techniques which aim is to classify points or objects to be analyzed into well separated groups or clusters. The existing k-means clustering methods discussed in this paper are the Forgy's method, Lloyd's method, MacQueen's method, and the Hartigan and Wong's method. It was observed that the modified k-means method performed relatively better than the enhanced k-means method and the other existing methods in terms of minimizing the total intra-cluster variance and accuracy using simulated data and real-life data sets.*

KEYWORDS: K-Means Clustering, Centroid Update, Euclidean Distance, Intra-Cluster Variance, Unsupervised Classification

INTRODUCTION

Cluster analysis is a multivariate technique where a set of data, usually multidimensional is classified into clusters (groups) such that members of one cluster are similar to one another with respect to some predetermined criterion (Anderberg, 1973; Hartigan, 1975; Jain and Dubes, 1988; Gan et al., 2007; Everitt et al., 2011; Yuan and Yang, 2019). The clusters of objects should exhibit high internal (within-clusters) homogeneity and high external (between-clusters) heterogeneity. Clustering is carried out on the basis of similarities or distances (Johnson and Wichern, 2002).

Cluster analysis as a field of study gained widespread acceptance in the sciences, and motivated world-wide research on clustering methods when Sokal and Sneath's (1963) publication of Principles of Numerical Taxonomy was published. The clustering method (cluster analysis) which is usually performed under a condition known as unsupervised learning is different from supervised classification (discriminant analysis). In supervised classification, observations are allocated to a known number of predefined groups, while in

unsupervised learning; neither the number of groups nor the groups themselves are known in advance.

Clustering methods can be broadly divided into two main groups which are based on the structure of their output namely: hierarchical and non-hierarchical clustering methods. Hierarchical clustering methods produce a sequence of the sets of the clusters. The clusters are merged (agglomerative methods) or split (divisive methods) step-by-step based on the applied similarity measure. The results of a hierarchical clustering method entail that agglomerative and divisive methods can be displayed graphically using a tree diagram known as dendrogram. While non-hierarchical or partitioning clustering methods partition the data, object set into clusters where every pair of object clusters is either distinct (non-overlapping) or has some members in common (overlapping). Partitioning clustering begins with a starting cluster partition which is iteratively improved until a locally optimal partition is reached. Amongst the partitioning clustering methods, the k-means method is the most popularly and commonly used in practice. K-means clustering is used to divide a set of objects (items, cases, entities, or data points) into k subsets or clusters (partitions, classes, or groups).

The purpose of this paper is to propose new k-means clustering methods that minimize the total intra-cluster variance, and also compare them with some existing k-means clustering method.

The rest of this paper is organized as follows: section 2 discusses the materials and methods from which the proposed methods are developed; section 3 is centered on experimental results and discussion, while section 4 is the conclusion of the paper.

METHODOLOGY

There are several k-means clustering methods that aim to classify data points to be analyzed into well separated clusters. Four existing k-means clustering methods were used namely: Forgy method; Lloyd's method; MacQueen's method; and Hartigan & Wong's method, considering the fact that Likas' method employs the MacQueen's method (or basic k-means algorithm) as a local search procedure, while Faber's method also adopted the MacQueen's method of updating centroids during initial partitioning (Oti and Onyeagu, 2020). The four-existing k-means clustering methods and the new proposed k-means methods have different centroid update approach and the rationale behind these developed methods is based on the assumption that an optimal clustering solution with k clusters can be obtained through local search. To be able to use any of these methods, the number of clusters present in the data set need to be known; multiple runs or trials will be necessary to find the best number of clusters (Oti and Onyeagu, 2020). There is no best method, as the tendency of generating global optimum depends on the characteristics of the data set (size, number of variables in the cases). The k-means clustering methods have two phases of iteration namely: the assignment or initialization phase which involves an iterative process where each data point is assigned to its nearest centroid using any metric of choice; the next is the centroid update phase, where clusters centroids are updated given the partition obtained by the previous phase. The iterative process stops when no data point change clusters or some maximum number of iterations is reached.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 42-54)

www.abjournals.org

Forgy's Method

The Forgy's method is a batch algorithm often called an offline centroid clustering model. Forgy (1965) proposed a method which is seldom referred to as traditional k-means algorithm. The algorithm is based on the minimization of the average squared Euclidean distance between the data points and the cluster's center known as centroid. A centroid is the center of a geometric object and it is seen as a generalization of the mean. Batch algorithm is an algorithm where a transformative step is applied to all data-point (case) at once, where c is the cluster center in the Euclidean distance and x is the case it is compared to, i is the dimension of $x(or\ c)$ being compared and k is the total number of dimensions. That is,

$$d_{euc} = \sqrt{\sum_{i=1}^{k}(c_i - x_i)^2} \quad (1)$$

being the most common distance.

Forgy's method starts with the choosing of k instance or initialization of data set uniformly at random and assigns the rest of the data points to the closest cluster (Peña et al., 1999). This method is very applicable because of its simplicity and high-speed intensity. It also treats the data set as a continuous distribution. Given the data set $\{x_1, x_2, \ldots, x_n\} \epsilon R^d$, where R^d is the real d-dimensional data space (or the Euclidean d-dimensional data space), the algorithm tries to find a set of k cluster centers $c = \{c_1, c_2, \ldots, c_k\} \epsilon R^d$. The error function for a continuous distribution is defined as

$$E = \sum_{i=1}^{k} \int f(x) d(c_i, x_i) dx \quad (2)$$

In the above equation, $f(x)$ is the probability density function at x and $d(c_i, x_i)$ is the distance function. We note that if the probability density function is not given (or known), then it has to be deduced (generated) from the given data. Though the k-means algorithm converges to a local optimum, the limit point depends on the initial points. Hence, it is appropriate to start with a reasonable initial partition in order to achieve high quality clustering solution. However, there is no efficient and universal technique for obtaining such initial partitions theoretically.

Algorithm 1: The Forgy's Algorithm.

1. Begin with any desired initial configuration. Go to step 2 if beginning with a set of seed points; go to step 3 if beginning with a partition of the data units.
2. Assign each data unit to the cluster with the nearest seed point. The seed points remain fixed for a full cycle through the entire data set.
3. Compute new seed points as the centroids of the cluster of the data units.
4. Repeat step 2 and 3 until the process converges; that is, continue until no data units changes their cluster membership at step 2.

Lloyd's Method

Lloyd (1982) proposed a method that is widely known as the standard k-means algorithm; it is also a batch algorithm that is based on the minimization of the average squared Euclidean

distance between the data items and the cluster centers like the Forgy's method. The dissimilarity between the Lloyd algorithm and the Forgy algorithm is that the Lloyd algorithm treats the data set as a discrete distribution while the Forgy algorithm treats the data set as a continuous distribution. While the similarity between them is that they have the same procedure. The error function for a discrete distribution is defined as

$$E = \sum_{i=1}^{k} \sum_{j=1}^{n} f(x) d(c_i, x_i) \quad (3)$$

In Equation (3) above, $d(c_i, x_i)$ is the distance function of the data point x_i and cluster center c_i. The first step of the algorithm begins with choosing the number of clusters k and its initial centroids or cluster centers. It could be done by either using k random observations or from the k observations that are the farthest from one another in the data space. Initialization of the centroids occurs only once, and once the initial centroids have been chosen; iterations are done on the following two steps. First, data set is assigned to cluster centroids (centers), using any of the distance metric. All cases assigned to a centroid are said to be part of the centroids subspace c ($R^{\mathbf{d}}$) (Morissette and Chartier, 2013). Second, update the value of the centroid by using the mean of the data points (cases) assigned to the centroid.

Algorithm 2: The Lloyd's Algorithm.

1. Choose k data objects representing the cluster centroids.

2. Assign each data object of the entire data set to the cluster having the closest centroid.

3. Compute new centroid for each cluster by averaging the data observations belonging to the cluster.

4. If at least one of the centroids has changed, go to step 2, otherwise go to step 5

5. Output the clusters.

Macqueen's Method

MacQueen (1967) proposed the MacQueen's algorithm, and it is often referred to as basic k-means algorithm, which is an online (or incremental) algorithm. The MacQueen's method is similar to the Forgy's and Lloyd's Methods, but the main difference is that the centroids are updated by re-calculating the points (cases) any time it is moved. Once the initial centroids have been chosen in the same way like the Lloyd's algorithm, the iterations follows: For each case (x_i) in turn, after arbitrarily partitioning of points (items) into clusters, we compute the coordinates ($\bar{x}_i'^s$) of the cluster centroid (mean), likewise the Euclidean distance is computed for each point from the group centroids and reassign each point to the nearest group. If a point is moved from its initial position, the cluster centroid must be recalculated or updated before computing the squared distances.

If the centroid of a case belongs to the nearest subspace, no change is made. If another centroid is closest to the subspace, the case is re-assigned to the other centroid and the centroids for both the old and new subspaces (centers) are recalculated as the mean of the cases. When we see that each point is currently assigned to the clusters with the nearest centroid, the process stops.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 42-54)

www.abjournals.org

Algorithm 3: The MacQueen's Algorithm.

1. Choose k points as initial cluster centroids.

2. Assign each object to the cluster that has the closest centroid.

3. When all objects have been assigned, re-compile the positions of the k centroids.

4. If at least there is a change in one of the centroids, repeat step 2 and 3, otherwise go to step 5.

5. Output result.

Hartigan and Wong's Method

Hartigan and Wong's method is a non-Lloyd heuristic that updates centers considering each point, rather than after each pass over the entire data set (Hartigan and Wong, 1979). Hartigan and Wong (1979) proposed the conventional k-means algorithm popularly known as Hartigan and Wong's algorithm.

It follows that the algorithm searches for the partition of data space with locally optimal within-cluster sum of squares error (SSE), which means that it may assign a case to another subspace, even if it currently belongs to the subspace of the closest centroid; doing so minimizes the total within-cluster sum of square (Morissette and Chartier, 2013). The initialization of the cluster centers is done in the same way as that of Lloyd's and Forgy's algorithm. The points (cases) are designated (assigned or allotted) to the centroid nearest to them and the centroids are calculated as the mean of the designated data points. The iterative steps are as follows:

Step 1. For each point I($I = 1, \dots, M$), find its closest and second closest cluster centers, $IC1(I)$ and $IC2(I)$, respectively. Assign point I to cluster $IC1(I)$.

Step 2. Update the cluster centers to be the average of the points contained within them.

Step 3. Initially, all clusters belong to the live set (specified number of k).

Step 4. This is the optimal-transfer (OPTRA) stage: Consider each point I ($I = 1, 2, \dots, M$) in turn. If cluster L ($L = 1, 2, \dots, K$) is updated in the last quick-transfer (QTRAN) stage, then the cluster belongs to the live set throughout this stage. Otherwise, at each step, it is not in the live set if it has not been updated in the last M optimal-transfer steps. Let point I be in cluster L1. If L1 is in the live set, do step 4a; otherwise, do step 4b.

Step 4a. Compute the minimum of the quantity, $R2 = [NC(L) * D(I,L)^2]/[NC(L) + 1]$, over all clusters $L(L \neq L1,\ L = 1,2, \dots, K)$ where the number of points in cluster L is denoted by $NC(L)$; while number of points in cluster $L1$ be $NC(L1)$; $D(I,L)$ is the Euclidean distance between point I and cluster L; $D[I, L(I)]$is the Euclidean distance between I and the cluster mean of the cluster containing I; $D(I,L)^2$ is the squared Euclidean distance between point I and cluster L. Let L2 be the cluster with the smallest R2. If this value is greater than or equal to $R1 = [NC(L1) * D(I,L1)^2]/[NC(L1) - 1]$, no reallocation is necessary and L2 is the new $IC2(I)$. Otherwise, point I is allocated to cluster L2, and L1 is the new IC2 (I). Cluster centers are updated to be the means of points assigned to them if reallocation has taken place.

The two clusters that are involved in the transfer of point I at this particular step are now in the live set.

Step 4b. This step is the same as step 4a, except that the minimum R2 is computed only over clusters in the live set.

Step 5. Stop if the live set is empty; otherwise, go to step 6; after one pass through the data set.

Step 6. This is the quick-transfer (QTRAN) Stage: Consider each point $I(I = 1, 2, \dots, M)$ in turn. Let $L1 = IC1(I)$ and $L2 = IC2(I)$. It is not necessary to check the point I if both the clusters $L1$ and $L2$ have not changed in the last M steps. Compute the values:

$R1 = [NC(L1) * D(I, L1)^2]/[NC(L1) - 1]$ and $R2 = [NC(L2) * D(I, L2)^2]/[NC(L2) + 1]$

If R1 is less than R2; point I remains in cluster $L1$. Otherwise, switch $IC1(I)$ and $IC2(I)$ and update the centers of clusters $L1$ and $L2$. The two clusters are also noted for their involvement in a transfer at this step.

Step 7. If no transfer took place in the last M steps, go to step 4, otherwise go to step 6.

Algorithm 4: The Hartigan and Wong's Algorithm.

1. Choose the number of clusters, k, and tentative centroids, $c_1, c_2, \dots, c_k$.

2. Observe an entity $i \in I$ coming either randomly or according to a pre-specified (dynamically) changing order.

3. d_{ij}= distance between case i and cluster j;

4. $d_{ij} = \arg min_{\ 1 \le j \le k}\, d_{ij}$

5. Assign cases *i* to cluster n_i ;

6. Re-compute the cluster means of any changed cluster above;

7. If no further change of cluster membership occurs in a complete iteration;

8. Output results.

The Modified K-Means Clustering Method

This first proposed method begins with first choosing the desired number of clusters k, its initial cluster partition, and the coordinates of the cluster initial centroid is denoted by $\bar{c}(i, j)$ which is the centroid of the jth variable over the data points in the ith cluster and it is computed as the arithmetic mean. This method also uses the squared Euclidean distance and the minimum distance rule like those of the existing k-means heuristic clustering methods by assigning each entity or data point to its closest (nearest) centroids. Specifically, for each data point i $\epsilon\ I$; its squared distances to the centroids are calculated. This method which proceeds in an incremental way (that is, adding cluster centers one by one as clusters are being formed) is such that when a case (point) is moved from the initial configuration, the cluster centroids

will be updated or recalculated before computing the squared distance. The ith coordinate, where $i = 1,2, \dots, k,$ of the centroid is updated using Equation (4) and Equation (5) below:

$$\bar{c}_i, new = \frac{N_k \bar{c}_i + c_{ij}}{N_k + 1} \quad (4)$$

if the jth point is added to the cluster.

$$\bar{c}_i, new = \frac{N_k \bar{c}_i - c_{ij}}{N_k - 1} \quad (5)$$

if the jth point is removed from the cluster.

Here N_k is said to be the number of points (cases) in the old cluster with centroid $\bar{c}^I = (\bar{c}_1, \bar{c}_2, \dots, \bar{c}_k)$ or perhaps the cluster size and centroid $\bar{c}_k$ is a multidimensional vector which minimizes the sum of squared distance to clusters elements. If a point or case is closest to the centroid of a particular subspace where the case is not moved to another cluster implies that the case will not be reassigned but if a case is closest to the centroid of a particular subspace where the case is moved to another cluster implies that the case will be reassigned and updated. The stopping rule is to end when there is no further change of cluster membership observed.

Algorithm 2.6: The Modified K-Means Algorithm.

1. Initial setting. Choose the number of clusters, k, and tentative centroids, $c_1, c_2, \dots, c_k$.

2. Apply minimum distance rule to determine what cluster list a data point, i, should be assigned to.

4. Update within cluster centroid, c_k, with Equation (4) or Equation (5) depending on if a point is added to a cluster or a point is removed from a cluster.

5. The stopping condition is to end when there is no further change of cluster membership observed.

6. Output results.

The Enhanced K-Means Clustering Method

This second propose method uses the Minkowski's distance, or r-metric, between vectors or N-dimensional points where $y = (y_v)$ and $c = (c_v)$ which is defined by the formula

$$d(y, c) = [\sum_{v=1}^{N} |y_v - c_v|^r]^{1/r} \quad (6)$$

In Equation (6), y_v are data points, c_v are cluster centers (centroids) and $\sum_{v=1}^{N} |x_v - y_v|^r$ is the r Minkowski distance. In application, when values $r = 2$ (Euclidean metric), $r = 1$ (Manhattan, or city block, metric) and $r \to \infty$ (Chebyshev, or Maximum, metric). However, the Euclidean k-means criterion is the usual k-means when $r = 2$ which is stated as

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 42-54)

www.abjournals.org

$$E = W(s,c) = \sum_{k=1}^{K} \sum_{i=S_K} d_{euc}^2(y_i, c_k)$$

where k represents the number of clusters, $c_k \epsilon\, c = \{c_1, c_2, \dots, c_k\}$ is the centroid of cluster s_k, $d_{euc}^2(y_i, c_k)$ is the squared Euclidean distance between an entity (cluster point) $y_i \epsilon\, s_k$ and its respective centroid c_k. The Minkowski k-means criterion allows the use of any distance function and W(s,c) is the square error criterion which is the sum of values over all clusters. Focusing on the Minkowski metric, which is between the N-dimensional entities y_i and c_k and is defined by

$$d(y_i, c_k) = [\sum_{V=1}^{N} |y_{iv} - c_{kv}|^r]^{1/r} \quad (7)$$

r is the exponent or power of Equation (7) which becomes

$$W_r(s,c) = \sum_{k=1}^{K} \sum_{i=S_K} d^r(y_r, c_k) = \sum_{k=1}^{K} \sum_{i \epsilon S_k} \sum_{v=1}^{N} |y_{iv} - c_{kv}|^r \quad (8)$$

This method is a batch k-means algorithm in which the minimum distance rule applies with the distance being the r power of Minkowski r-metric rather than the squared Euclidean distance (Amorim and Komisarczuk, 2012; Amorim, 2012; Amorim and Mirkin, 2012).

Algorithm 6: The Enhanced K-Means Algorithm.

1. Choose at random the number of cluster centers (centroids) $c = c_1, c_2, \dots, c_k$.
2. Calculate the distance between each data point and cluster centers using Equation (7)
3. Assign data point to the cluster center whose distance from the cluster center is the minimum of all cluster centers.
4. New cluster center is calculated using $v_i = \frac{1}{|c_i|} \sum_{y \epsilon c_i} y_i$ where $|c_i|$ denotes the absolute value of data points in ith cluster and v_i is the mean of the cluster c_i and $\sum y_i$ is the sum of points or cases in the data space.
5. The distance between each data point and new obtained cluster centers is recalculated.
6. If no data point was reassigned then stop, otherwise repeat step 3 to 5.

RESULTS AND DISCUSSION

This section shows the performance and also the comparison of the new k-means clustering methods modified and some of the existing k-means clustering method using R statistical software (R version 3.2.2) support window 64-bit system. We conducted experiments using one simulated data set and two real-life data sets to ensure the efficiency of the proposed methods. The number of clusters k used is two and three, since research has proven that the optimal number of clusters k will either be two, three, or four using methods like elbow, the silhouette and the gap statistic methods (Kaufman and Rousseeuw, 1990). The performance of the proposed methods was evaluated using total intra-cluster variance and accuracy parameters.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 5, 2020 (pp. 42-54)

www.abjournals.org

Total intra-cluster variance: The total intra-cluster variance is defined as the sum of squared distance between points and the corresponding centroid. That is; $W(C_K) = \sum_{x_i \epsilon c_k}(x_i - \mu_k)^2$ where

- x_i is the data point belonging to the cluster c_k.
- μ_k is the mean value of the points assigned to the cluster c_k.

Accuracy: Accuracy is defined as the ratio of the total number of correctly classified instances divided by total number of correctly plus incorrectly classified instances.

Simulated Data

The simulated data was generated randomly from a Gaussian (Normal) distribution with dimension of 250 rows and 2 columns (categories or attributes) that are divided into two and three clusters (that is, k = 2, 3). We chose 300 true centers uniformly at random given the above dimension. The point from the Gaussian distributions has a variance of 1 around each true center. Thus, this obtained a number of well separated Gaussians with the true centers providing a good approximation to the optimal clustering. The Gaussian distribution is used in the simulation of data because it is suitable for most applications and it's also the most commonly used distribution in statistics owing to the fact that it has finite moments (mean, variance …) for small parameter values.

Shown below is the summary table of the results of experiments and data analysis of some of the existing method when the number of clusters k is two and three respectively:

Table 1: Summary table of the Gaussian simulation when the number of clusters k = 2 and 3

Methods	***when the number of clusters k* = 2**			***when the number of clusters k* = 3**		
	Mean	**Standard Deviation**	**Accuracy in (%)**	**Mean**	**Standard Deviation**	**Accuracy in (%)**
Forgy	1.584	0.4949	80.0	2.248	0.7476	83.7
Lloyd	1.496	0.5020	79.1	1.920	0.8092	79.0
MacQueen	1.504	0.5020	79.1	2.296	0.7831	81.4
Hartigan & Wong	1.504	0.5020	79.1	2.144	0.8299	78.3
Modified k-means	1.880	0.3263	86.8	2.295	0.6898	86.5
Enhanced k-means	1.601	0.4604	81.8	2.234	0.7131	84.3

From the above results of the simulation generated randomly, when the number of clusters $k = 2$, the modified k-means method performed better than the enhanced k-means method and other existing k-means clustering method with minimum standard deviation of 0.3263 and high accuracy of 86.8 percent, considering the fact that the variance (the total within-cluster sum of squares) is minimized; it measures the compactness (i.e. goodness) of the clustering which is meant to be as small as possible, also, high accuracy indicates how better

the method is and it is expressed in percentage. The number of clusters $k = 3$, the modified k-means method also performed best with a standard deviation of 0.6898 and accuracy of 86.5 percent.

Real-Life Data

To understand how efficient these methods are under more practical circumstances, we run a number of experiments on two data sets which consist of the iris data set, and the wine data set. The data sets are both from UC-Irvine Machine Learning Repository. Each experiment involves solving k-means problem on a set of points in a real dimensional space.

Iris Data Set

The iris flower data set is a multivariate data set with 150 rows (instances) which is divided into 3 instances each, where each class refers to a type of iris plant (iris setosa, iris versicolor, and iris virginica): the number of attributes is 4 which consist of sepal length, sepal width, petal length and petal width (Fisher, 1936). The summary table of the results when the number of clusters k is two and three is shown in the Table 2 below:

Table 2: Summary results of iris data when the number of clusters k = 2 and 3.

Methods	***when the number of clusters k* = 2**			***when the number of clusters k* = 3**		
	Mean	**Standard Deviation**	**Accuracy in (%)**	**Mean**	**Standard Deviation**	**Accuracy in (%)**
Forgy	1.3533	0.4796	83.50	1.560	0.8067	82.00
Lloyd	1.6467	0.4796	83.50	2.4933	0.7396	85.20
MacQueen	1.3533	0.4796	83.50	1.9333	0.5983	91.50
Hartigan & Wong	1.6467	0.4796	83.50	2.080	0.8633	79.10
Modified k-means	1.8014	0.3148	90.29	1.8622	0.6765	87.68
Enhanced k-means	1.7823	0.3325	89.70	1.9467	0.8035	82.45

From the above experiments and summary table on iris data set, it is observed that when the number of clusters k = 2, the modified k-means method performed better than the enhanced k-means method and also the other existing methods with standard deviation of 0.3148 and accuracy rate of 90.29 percent. Also, when the number of clusters k = 3, the MacQueen's method performed better than the proposed methods and every other existing methods with standard deviation of 0.5983 and 91.50 percent accuracy; the modified k-means method performed better than the enhanced k-means method, Forgy's method, Lloyd's method and Hartigan & Wong's method with minimum standard deviation of 0.6765 and 87.68 percent accuracy.

Wine Data Set

The wine data set is a multivariate data with 178 numbers of rows (instances) and three classes with 13 attributes (columns). The attributes of data set are alcohol, malic acid, ash,

alkalinity of ash, magnesium, phenols, flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue, 0D280/0D315 of diluted wines and proline. The output of the experiments when the number of clusters k = 2 and 3 will be summarized in Table 3 below:

Table 3: Summary results of wine data when the number of clusters k = 2 and 3.

Methods	***when the number of clusters k* = 2**			***when the number of clusters k* = 3**		
	Mean	**Standard Deviation**	**Accuracy in (%)**	**Mean**	**Standard Deviation**	**Accuracy in (%)**
Forgy	1.0785	0.2692	91.45	2.5946	0.7476	83.7
Lloyd	1.9209	0.2689	91.80	2.1763	0.8092	79.0
MacQueen	1.0785	0.2689	91.80	2.1307	0.7831	81.4
Hartigan & Wong	1.9215	0.2689	91.80	2.2389	0.8299	78.3
Modified k-means	1.9190	0.2673	91.96	2.1805	0.4923	85.27
Enhanced k-means	1.9125	0.2655	92.15	2.4489	0.6282	76.20

From the above summary table on wine data set, it was observed that when the number of clusters k = 2, the enhanced k-means method performed better than the modified k-means method and other methods with a minimal standard deviation of 0.2655 and accuracy of 92.15 percent. When the number of clusters k = 3, the MacQueen's method outperformed every other method with standard deviation of 0.4310 and accuracy of 88.10 percent. The performance of the modified k-means method was relatively efficient than the enhanced k-means method, Forgy's method and Hartigan and Wong,s method with standard deviation of 0.4923 and accuracy of 85.27 percent.

CONCLUSION

In this paper, we have presented a modified k-means method that updates its clusters centroids depending on if a point is added to the cluster or if a point is removed from the cluster and also the enhanced k-means clustering method that uses the Minkowski's distance in calculating between each data points and the cluster centroids which yielded excellent results in terms of minimizing the total intra-cluster variance, and it was also shown to be more accurate than a variety of other methods while comparing its performance with them. From experimental results, the modified k-means method outperformed the enhanced k-means method and other existing methods in the simulated data and also in the wine data when the number of clusters $k = 2\ and\ 3$; the enhanced k-means method performed better than other methods in the wine data when the number of clusters $k = 2$, while the MacQueen's method outperformed the other methods when the number of clusters $k = 3$.

Our future research will be considering these methods with respect to their iterative time complexity using the personal computer time.

Acknowledgements

The authors wish to thank the editor and referees for their worthwhile comments and suggestions. This research was sponsored in part by Nigerian Tertiary Education Trust Fund (TETFUND).

REFERENCES

Amorim, R. C. (2012). Constrained clustering with Minkowski weighted k-means. Proceedings of the 13th IEEE International Symposium on Computational Intelligence and Informatics, 13-17.

Amorim, R. C., Komisarczuk, P. (2012). On Initializations for the Minkowski weighted k-means. International Symposium on Intelligent Data Analysis, 45-55.

Amorim, R. C., Mirkin, B. (2012). Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. Pattern Recognition, 45 (3), 1061-1075.

Anderberg, M. R. (1973). Cluster Analysis for Applications. New York: Academic Press.

Everitt, B., Landau, S., Leese, M., Stajl, D. (2011). Cluster Analysis, 5thedition, John Wiley and Sons.

Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems, "Annals of Eugenics, 3, 179-188.

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classification. Biometrics, 21, 768-769.

Gan, G., Ma, C., Wu, J. (2007). Data Clustering: Theory, Algorithms, and Applications, SIAM Series.

Hartigan, J. A. (1975). Clustering Algorithms: New York: John Wiley and Sons.

Hartigan, J. A., Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28 (1), 100-108.

Jain, A. K. and Dubes, R. (1988). Algorithm for Clustering Data: Eaglewood Cliffs Prentice Hall.

Johnson, R. A., Wichern, D. W. (2002). Applied Multivariate Statistical Analysis: 5th Edition, Eaglewood Cliffs, NJ: Prentice-Hall.

Kaufman, L., Rousseeuw, P. J. (1990). Finding Groups in Data, An Introduction to Cluster Analysis. Wiley Series, New York: John Wiley and Sons.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transaction on Information Theory, 28 (2), 129-137.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, (1), 281-297. Berkeley, CA: University of California Press.

Morissette, L., Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. Tutorials in Quantitative Methods for Psychology, 9 (1), 15-24.

Oti, E. U. and Onyeagu, S. I. (2020). Some versions of k-means clustering method and its comparative study in low and high dimensional data: African Journal of Mathematics and Statistics Studies, 3 (1), 68-78.

Peńa, J., Lozano, J. and Larrańaga, P. (1999). An empirical comparison of four initialization methods for the k-means algorithm: Pattern Recognition Letters, 20 (10), 1027-1040.
Sokal, R. and Sneath, P. (1963). Principles of Numerical Taxonomy: San Francisco, California.
Yuan, C. and Yang, H. (2019). Research on k-value selection method of k-means clustering algorithm: Multidisciplinary Scientific Journal, 2 (2), 226-235.