ENHANCED CENTROIDS INITIALIZATION OF K-MEANS ALGORITHM FOR FIXED ASSET MONITORING

ALETA C. FABREGAS

Polytechnic University of the Philippines

Abstract: This research developed a new method of initializing centroids for K-means algorithm and applied in determining the status of the fixed asset. The new method introduced the comparator function to determine the appropriate initial centroids and proved to be more efficient than the standard K-means. Since the initial seeds of the traditional K-means algorithm are chosen randomly leading to more looping and computational time, the improved method is proposed to eliminate the unspecified selection of the initial cluster and reduce inconsistency in the final clustering.

The major responsibility of the Asset Management Office (AMO) is to determine the maintenance level of various equipment of the Polytechnic University of the Philippines (PUP). However, with a lot of equipment it managed, some are not properly maintained and this resulted to a very costly repair. In order to implement the proper repair to the equipment, the age factor based on the depreciation and usage factor of the physical asset are utilized in the Enhanced Initialization of Centroids for K-means algorithm. The clustering result of the new method produced more consistent level of maintenance. Once the equipment is clustered using the new method, that same equipment is no longer clustered into another level of maintenance unlike the standard K-means, and this is how the author defined duplicates. Local optimization or convergence of clustering is faster in the new method as compared to the old K-means algorithm.

This has been developed to solve the problem of AMO, knowing the status of the equipment based on age and usage factors to determine what is the proper level of maintenance (major or minor repair) with the higher percentage of consistency in the clustering of equipment is needed in the fixed asset monitoring. The enhanced centroids initialization of K-means is adapted by the AMO of the PUP in Sta. Mesa, Manila.

The fixed asset monitoring application was developed using Java Eclipse obtaining actual data from a database using Mysql. To test the proposed algorithm, the fixed asset monitoring system also implemented the standard initialization of K-means algorithm for comparison. The test yielded more favorable results using the enhanced centroid initialization using new method based on speed and reduced number of duplicates, making the new method of fixed assets monitoring application more efficient and consistent. The data used are filtered and finalized by AMO. These are the following: (1) acknowledgement receipt for equipment for the years 2009, 2012, and 2014 with the acquisition date and acquisition cost of the ICT equipment; (2) office equipment and the inventory report of motor vehicle in 2016 with the description of utility vehicle containing the acquisition cost, date and status of running condition.

Keywords: *K-means algorithm, comparator function, depreciation, age factor, usage factor*

1. INTRODUCTION

K-means clustering is a process of cluster analysis which aims to divide *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean (Da Costa, 2013). This algorithm started by selecting k points as the initial k cluster seeds and obtaining these initial seeds include random selection from the dataset and the method repeats between two steps until reaching stopping point. The partitioning method of K-means algorithm is used in the study. Most partitioning methods are distance-based. Given k, the number of partitions to construct a partitioning method creates an initial partitioning and then uses an iterative relocation technique that attempts to improve the partitioning by moving objects from one group to another. In a good partitioning, the objects in the same cluster are close or related to each other, whereas, objects in different clusters are far apart or different (Puri, 2015). This study formulated a new seed initialization method of the K-means algorithm obtaining the highest and lowest pairs of values from the two attributes (x and y coordinates) as guide to input the values in Euclidean distance with two dimensions with the implementation of the new method presented, the clean boundary of the clustering between points and overlapping between the groups is reduced immediately.

The fixed assets, that included property, plant and equipment, are tangible assets held by an entity for the production or supply of goods and services, for rentals, and for administrative purposes. These assets are expected to be used for more than one accounting period. Some of the fixed assets are buildings, land, furniture and fixtures, machines and vehicles. Depreciation of fixed assets is needed in accounting for two reasons. The first is that the asset is maturing to the end of its useful life. The older the asset, the fewer years it has left to produce income. Second, is that due to wear and tear, older assets decline in productivity and would require higher repair and maintenance costs.

The Property, Plant and Equipment (PPE) in the government agencies were applying prescribed depreciation method by Commission on Audit (COA). But recently government accounting is modified and specified at the latest Circular Number 2015-007 dated October 22, 2015, which is the Government Accounting Manual for National Government Agencies. The Polytechnic University of the Philippines (PUP), is the main client of the study. One of the most common problems in the university is the application of repair and maintenance of fixed assets from the time of acquisition to the point of disposal. One way of managing the fixed assets is to monitor the depreciation. The allocated cost is referred to as depreciation. The university owns large number of equipment for each type of fixed asset and the acquisition date of each equipment varies. The acquisition date of the various equipment result in the difficulty of preparing financial report on- time and maintaining fixed assets properly. To solve such problem, the K-means algorithm will help address the tediousness of preparing the financial report of fixed assets in relationship to depreciation and determining the level of maintenance categorized into two; major and minor repairs (this is adapted from Section 24–Repairs and Maintenance of Government Accounting Manual).

The Asset Management Office (AMO) with its primary function of safeguarding the property of the university also monitors numerous types of fixed assets or PPE. It also recommends the approval of the repair on condemnation of various equipment from different units of the University. The AMO consolidates all reports of the fixed assets. The fixed assets monitoring is a tedious and crucial job for the AMO. The AMO as the major repository of all the types of current and fixed assets of the university and the manager of all assets is also the source of all reports submitted to the Accounting Department (AD) for reconciliation purposes. It is in this light that the study about new procedure was implemented to check the status of the fixed asset based on acquisition cost and lifespan and the output is the formulation of new ways of checking the fixed asset.

This study improved the K-means algorithm by incorporating the Enhanced Centroid Initialization at fixed asset monitoring system to determine the level of maintenance. Specifically, it aimed to achieve the following: (1) to develop an enhanced centroid initialization technique that will implement the comparator function on a dataset to get the highest and lowest pair of values from the dataset; (2) to simulate the use of Enhanced K-means Algorithm in terms of clustering results; and (3) to measure the improvement of the Enhanced K-means vis-à-vis Standard K-means in terms of clustering consistency.

For the stakeholders of this project: the Director and staff of the AMO, AD, the Information and Communication and Technology Center (ICTC), motorpool, and the Community Development Management Office (CDMO) of the PUP will be the beneficiaries. The AMO and the AD would no longer spend long hours preparing the financial reports that include depreciation reports and application of level of maintenance. On the other hand, ICT, motorpool and CDMO department will be properly implemented in the fixed assets. The proper implementation of Enhanced Centroid Initialization of K-means Algorithm (ECIk) will help the administration in safeguarding the fixed assets without spending unreasonable cost for corrective maintenance.

In a previous paper (Fabregas *et al.,* 2016), the new algorithm was compared with Standard K-means algorithm with random selection of initial centroids. The previous test yielded positive results showing that the ECIk algorithm is better than the original in terms of execution speed, reliability and consistency. In this study ECIk is implemented in the Fixed Asset Monitoring System (FAMS). The software and actual data from AMO were used to test the performance of the ECIk. The reports of Depreciation schedule and level of maintenance which is based on lifespan and acquisition cost were generated with higher execution speed, reliability and consistency using ECIk.

The procedure developed by MacQueen (1967) divided a sample of n entities into k sets based on a Euclidean distance measure. The algorithm assigned each item to the cluster having the nearest centroid (mean). The steps are: (1) partition the items into k initial clusters; (2) from the list of dataset, assigning an item to the cluster whose centroid (mean) is nearest (distance is usually computed using Euclidean measure). Recomputed the seed for the cluster receiving the new item and for the cluster losing the item; and (3) repeat step 2 until no more reassignments take place. Instead of starting with a partition of all items into k preliminary groups in step 1, one could specify k initial seed points and then proceed to step 2. The objects are partitioned such that they are as close to each other as possible within each cluster, and far from objects in other cluster (Das, 2003).

There are significant number of researches using k-means clustering algorithm managing current assets and fixed assets. Since K-means is applicable in large databases, and suited in numerical data, monitoring of fixed assets to determine the minor and major repairs will yield better results using this algorithm.

Aggarwal *et al.* (2012) solved the first limitation of the K-means which is the random selection of initial centroid by applying the distance function with some complexities in the process. The research compares the basic K-means and enhanced K-means algorithm and proves that enhanced K-means is more efficient. Rauf *et al.* (2012), proposed that the K-mean algorithm in which the initial seeds are computed and as the data is similar, it results in same calculations, so the number of iterations remains constant and the elapsed time is also improved. Raj & Punithavalli (2013), proposed a system named Median Unique Vector Optimization Algorithm that sorts out the correct selection of initial cluster centers for K-means which is possibly used to avoid the local optimum problem and may lessen the number of iterations after the clustering process. Dhanachandra *et al.* (2015), have also proposed a subtractive clustering method wherein it generates the centroid based on the potential value of the data points

In the study of Kaur & Dyoti (2013), K- means algorithm was proposed to solve the problem of dead unit and to optimize the selection of initial centroids of clusters by using most populated area as a centroid of cluster. It ensures the minimum execution time during the allocation of data points to respective clusters. There were two major steps made in their enhancement: (1) selection of initial centroids and (2) allocation of data points to respective clusters. These steps are similar to the suggestion of a Mathematics Professor from PUP in finding out the value of initial centroid of clusters by calculating mean of selected group and find out the distance matrix for clusters by calculating distance between centroids.

The study was challenged to develop the method that will not be using complex mathematical computations in the initial selection of centroids. But the enhancement applied the Comparator Interface method of Java programming language with the purpose of comparing two objects using the logical And operator to determine the highest and lowest pair of points. The result of the comparison served as the input values to the Euclidean distance formula of K-means algorithm. The comparator method in the initialization stage resulted to efficient clustering of the objects with faster convergence and minimum duplicates, thus, gave birth to the new method called the Enhanced Centroid Initialization of K-means algorithm (ECIk).

Currently, the AD and AMO of the university are using MS Excel in recording the inventory report that reflects the status of equipment and the depreciation report following the COA's standard accounting procedure stated in the Government Accounting Manual (GAM). The developed Fixed Asset Monitoring System (FAMS) with the ECIk will help the AMO, ICT, motor pool, CDMO and AD in preparing the depreciation schedule and generating reports of the level of maintenance, either major or minor repair to fixed asset. The depreciation factors used in the study are the age and usage factors. The age factor is the lifespan of the fixed asset that declines because of the systematic recording of the depreciation and the usage factor is the physical condition caused by the depreciation. The depreciation concept applied in the study used the principle of the Manual on Disposal of Government Property in the Philippines. If the fixed asset is not used, the Usage Factor (UF) is 0.90, otherwise, it is equal to the value assigned to the condition factor stated on Section 24 of GAM and this is one of the major result of clustering after using the algorithm.

The framework is a tool used to analyze and organize ideas. It shows the parameters and variables, the processes, and the expected outcome of the study. Figure 1 illustrates the scheduling agent or the software. The software requires a database to store the names of the fixed asset, the types of the fixed asset, the sub-categories, cost of acquisition, date of acquisition, service life and salvage value. To develop the system, Java (Eclipse Neon) is used while applying the K-means algorithm utilizing the enhanced initialization centroid. After the process, a depreciation schedule and level of maintenance into minor repair and major repair are automatically generated. This old K- means algorithm and the ECIk partition the datasets in such a way that items in the same cluster are more identical to each other than to those in other groups. The non-overlapping and non-subordinated clusters traits of K-means are essential in determining which of the fixed assets will be having minor and major repairs to maintain them properly.

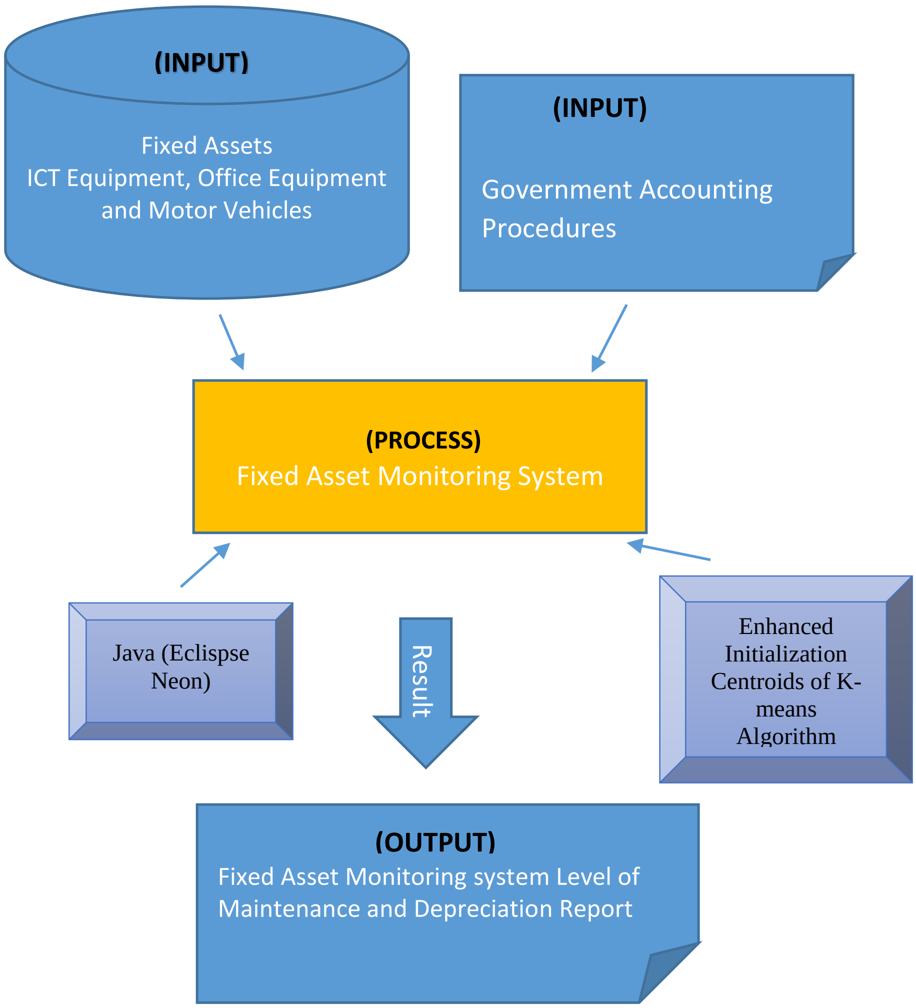


Figure 1. Conceptual framework of the study.

Figure 2 presented the fixed assets classified into ICT equipment, office equipment and motor vehicles, and the accounting guidelines and procedure for Depreciation as input to the process that are needed to generate the required reports. The process integrates the application of FAMS with the ECIk algorithm using Java Eclipse Neon. The process will use the age factor and usage factor of the fixed asset to be used in the system as input to the K-means algorithm and enhanced centroids initialization. The output resulted from the generation of depreciation schedule and the level of maintenance to be applied at the fixed asset.

2. METHODOLOGY

2.1 Software

The software is a stand-alone desktop application which basically stores the data in a local Mysql database. The application itself is developed using Java (Eclipse Neon version 1) so the program will run on a Windows platform particularly Windows 10. The program does not require a network or an internet connection for it to function.

2.2 Hardware

The system is developed and tested on an Intel(R) Core I 7 65000U CPU @ 2.50 GHz 2.60 GHz based laptop processor running at 2.50 GHz with 8.0 of RAM. It is currently running Windows 10, 64- bit operating system.

2.3 Data

The data used in the study are the documents coming from the PUP and AMO. These are the following: Acknowledgement Receipt for Equipment for the years 2009, 2012, and 2014 with the acquisition date and acquisition cost of the ICT equipment and Office Equipment and the Inventory report of Motor Vehicle in 2016 with Description of Utility Vehicle containing the Acquisition cost and date and status of running condition.

2.4 Experimental method

The experimental method is utilized by the author to simulate the actual data using the enhanced algorithm. The Rapid Application Development is adopted by the study to complete the whole system. It is a type of incremental model developed in parallel as if they were mini projects. The developments are time bounded, delivered and then assembled into a working prototype. This is suitable for containing such a project requiring shorter development times (tutorialspoint.com, 2017). The study used this model in order to determine if the objective of the study to the Enhancement of K-means algorithm is fitted in the FAMS application.

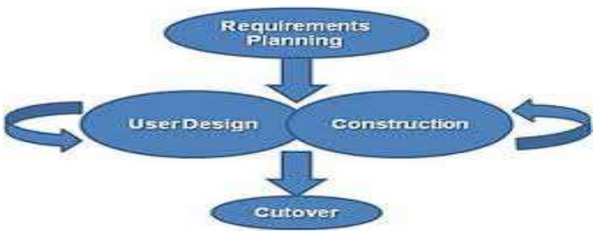


Figure 2. Rapid application development model.

2.5 K-means clustering method

The study used K-means algorithms to solve the depreciation scheduling and fixed asset monitoring problems. In this section, the study will discuss the K-means algorithm itself in a general point of view. Figure 3 shows the process flow of the K-means algorithm.

The K-means algorithm aims at minimizing the objective function known as squared error function called Euclidean distance.

$$J = \sum_{j-1}^{k} \sum_{i-1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (1)$$

where $\left\| x_i^{(j)} - c_j \right\|$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster c_j is an indicator of the distance of the *n* data points from their respective cluster centers.

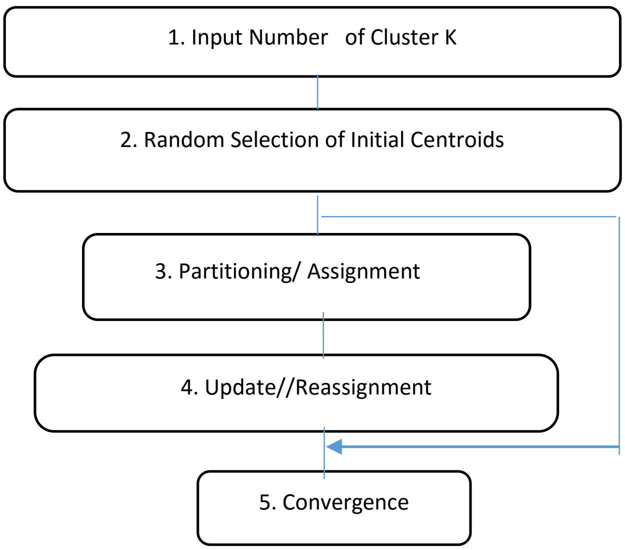


Figure 3. Flowchart of K-means algorithm.

The partitioning and update process used the Euclidean distance to find new centroids until the computed centroids reached convergence point. After analyzing the procedures of the standard K-means clustering algorithm, the study developed other method of initializing seeds in order to improve the performance of the algorithm by eliminating the random selection of the cluster center. The clustering result of the k-mean clustering algorithm lies greatly upon the reliability of the initial seeds which are selected randomly. The immediate seeds are without computational basis, which is leading to the less reliable output (Fabregas *et al.,* 2016).

2.6 The Enhanced Centroid Initialization of K-means algorithm

The Enhanced Centroid Initialization of K-means (ECIk) algorithm integrates the use of comparator function in obtaining the highest and lowest pairs of values serving as guide to determine the initial centroids. This method solves the limitation of the standard K-means algorithm while randomly selecting the initial centroids that produces less reliable result which was proven by the study.

The steps of the ECIk are the following:

1. Initialization.

    1.1 Setting the number K of clusters, obtain the highest and lowest pair of values of the object using the comparator method.

    1.2 The highest and lowest sets of values for each object is obtained by comparing them using comparator method. The resulting highest and lowest set of values will be the guide to select the initial seeds to create the initial partition.

2. Assignment/ Update.

    2.1 Assignment step, where the distances of the object from the centroids of each of K clusters are computed using the Euclidean distance; if the object is not currently in the cluster with the closest prototype, then it is reassigned to its nearest cluster.

    2.2 Update step: if reassignment occurs, both the clusters (gaining the new object and losing the old object) are updated and their seeds are recomputed using the current clustering;

3. Convergence. When the refinement steps of Assignment and Update Stopped reaching the local optimization.

The compare method shown in Figure 3 is used by the study to compare the sets of values (x = age factor, y = usage factor) from the dataset.

2.7 Simulation and evaluation

Figures 4 and 5 show the simple simulation of the traditional K-means algorithm as compared to ECIk using java program. Based on the input, the age factor (AF) and usage factor (UF) of the seven service vehicles contain values in the same behavior. The pair of values from the first to the fourth set is in increasing order and the pair from the fifth to seventh set is in decreasing order. From the concept of depreciation, when the pair of values are low, the service vehicle is old and the physical condition is low. After reaching the limit, based on the output, the result of the clustering of the service vehicles using the traditional algorithm of K-means with the two (2) attributes is: the first two Service vehicles are old and their physical conditions are not good, then it needs the type level (1-major) of maintenance to keep it running efficiently. And the next five (5) Service vehicles need type level (2-minor) of maintenance. The number of iterations is four (4) before reaching the convergence point

```
lowest of X  ← X[0];
lowest of Y ← f Y[0];
centroid1 ← [ X {0}, Y[0]] ;
….
FOR i ← 1;
 IF lowest of X > X[i] AND lowest of Y>  lowest of Y{I};
               THEN
                 lowest of X ←  X{I];
                 lowest of Y← Y[I];;
                  indexOfLowest ← I;
                 centroid1←[X[indexOfLowest], Y[indexOfLowest]];
 repeat Until I = size of items
 Highest of X ← X[0];
highest of Y Y[0];
centroid2 ← [X[0], Y[0]];
FOR i ← 1;
 IF highest of X < X[i] AND highest of Y <  Y[i]
                 THEN
                  Highest of X  ←  X[i];
                  Highest of Y ←  Y[i];
                 indexOfHighest ← I;
                 centroid2←Xi]ndexOfHighest, Y[indexOfHighest]];
repeat until I =size of items
```

Figure 3. The algorithm of initialization with comparator method of the enhanced centroid initialization.

Input	Output
Age factor: [1.0, 1.5, 3.0, 5.0, 3.5, 4.5, 3.5] Usage factor: [1.0, 2.0, 4.0, 7.0, 5.0, 5.0, 4.4]	Seed1: [3.0, 4.0]vehicle: 3 Seed2: [4.5, 5.0]vehicle: 6 Distance to C1: [3.6055512, 2.5, 0.0, 3.6055512, 1.118034, 1.8027756, 0.6403125] Distance to C2: [5.315073, 4.2426405, 1.8027756, 2.0615528, 1.0, 0.0, 1.1661904] Clustering 1: [1, 1, 1, 2, 2, 2, 1] Average of cluster 1: [2.25, 2.85] Average of cluster 2: [4.3333335, 5.6666665] Distance to C1: [2.2327113, 1.1335783, 1.372953, 4.9784536, 2.4869661, 3.1120734, 1.991231] Distance to C2: [5.7348833, 4.633813, 2.1343746, 1.490712, 1.0671874, 0.6871841, 1.5162086] Average of cluster 1: [2.25, 2.85] Average of cluster 2: [4.3333335, 5.6666665] Cluster of iteration 2: [1, 1, 1, 2, 2, 2, 2] Another Iteration Distance to C1: [1.5723301, 0.4714045, 2.034426, 5.639642, 3.1446605, 3.7712362, 2.6549745] Distance to C2: [5.3561296, 4.2559514, 1.7573061, 1.8676523, 0.71632737, 0.51295704, 1.1371564] Average of cluster 1: [1.8333334, 2.3333333] Average of cluster 2: [4.125, 5.35] cluster of iteration 3: [1, 1, 2, 2, 2, 2, 2] Another Iteration Distance to C1: [0.559017, 0.559017, 3.0516388, 6.6567636, 4.160829, 4.776243, 3.6704905] Distance to C2: [5.0056367, 3.9046638, 1.4058449, 2.2127812, 0.40792164, 0.6053097, 0.7889232] Average of cluster 1: [1.25, 1.5] Average of cluster 2: [3.9, 5.08] Done **Cluster of iteration 4: [1, 1, 2, 2, 2, 2, 2]** **Iteration Done: 4**

Figure 4. The output of the K-means algorithm.

Input	Output
Age factor: [1.0, 1.5, 3.0, 5.0, 3.5, 4.5, 3.5] Usage factor: [1.0, 2.0, 4.0, 7.0, 5.0, 5.0, 4.4]	seed1: [1.0, 1.0]vehicle: 1 seed2: [3.5, 4.4]vehicle: 7 distance to C1: [0.0, 1.118034, 3.6055512, 7.2111025, 4.7169905, 5.315073, 4.2201896] distance to C2: [4.2201896, 3.1241, 0.6403125, 3.001666, 0.5999999, 1.1661904, 0.0] clustering 1: [1, 1, 2, 2, 2, 2, 2] average of cluster 1: [1.25, 1.5] average of cluster 2: [3.9, 5.08] Done **cluster of iteration 2: [1, 1, 2, 2, 2, 2, 2]** **Iteration Done: 2**

Figure 5. The output of the improved initialization of centroids of the K-means algorithm.

Using the same set of age and usage factor, the result of integrating the comparator method in the enhancement is: the seed 1 is the vehicle 1 with [1.0, 1.0] and seed 2 is the vehicle 4 with [5.0, 7.0]. Applying the algorithm of the improved initialization of centroids resulted to the application of level 1 of maintenance in the first two (2) service vehicles and the level 2 in the next five (5) vehicles. The number of iterations for the improved K-means algorithm is reduced to two (2) as compared with the standard K-means with four iterations. The convergence step is easier to achieve in the improved algorithm.

With this experiment, the author was able to determine the number of factors affecting the performance of the Enhanced Centroid Initialization as compared to the traditional K-means algorithm. The following are: (1) Speed and Iterations; (2) Number of Duplicates; and (3) Percentage of Consistency. The Speed of clustering of the two (2) algorithms are compared using milliseconds. The iterations is the number of times a certain equipment is clustered to reach local optimum or convergence. The number of duplicates is based on how many times a certain equipment is clustered into minor or major repair with the ECIk and K-means algorithm (KMA). And the percentage of consistency is based on the number of duplicates. Consistency Percentage = ((Number of Assets – Count of Assets with Two Type cluster)/ (Number of Assets)) * 100, to determine how consistent in terms of duplicates is the two algorithms.

3. RESULTS AND DISCUSSION

The two (2) algorithms are evaluated by the system and presented in the Dashboard module. The consistency constraint is based on the number of iterations for each classification of the fixed asset. ECIk has fewer number of iterations because the convergence step is immediately achieved as compared with the KMA. The clustering button on the Clustering Module for the two algorithms could be clicked several times. Every time the clustering button is pressed, the number of iterations for the KMA is changed, unlike in the ECIk. And this resulted to the consistency of the ECIk. Thus, this is shown in the Queries module. Clustering of the fixed asset based on KMA made the result inconsistent because an asset could be grouped in the other set, as the cluster button is pressed.

The speed constraint for the two algorithms are not consistent. In different situations of the process, the KMA is faster than the ECIk. If the number of the dataset is few, the ECIk is faster than the KMA, but KMA is faster in the case of increasing dataset.

Figure 6 presented how the Clustering Module does the Clustering Process. This module clusters the fixed assets using the ECIk and KMA algorithms and resulted in two levels of maintenance, major and minor repairs for individual, subcategory and all categories of fixed assets are presented in the Queries Menu.

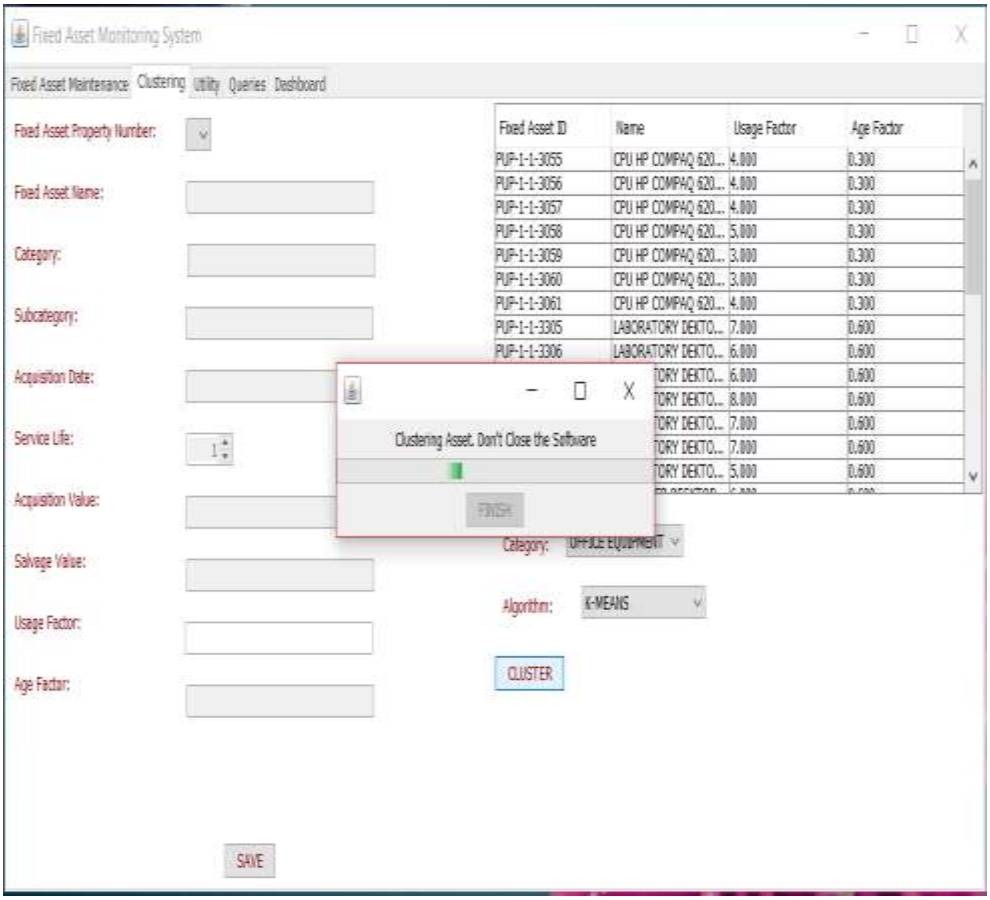
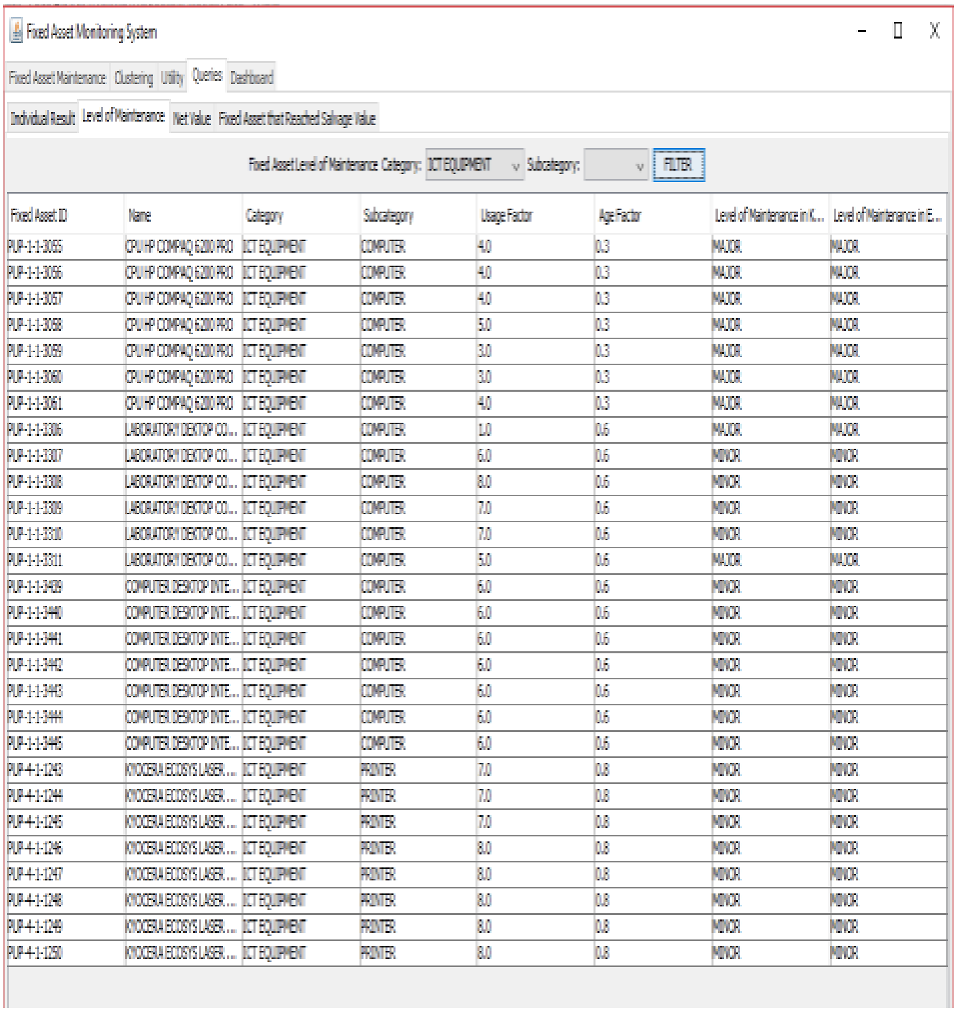


Figure 6. Screenshot of the Clustering Process for Fixed Asset Monitoring System.

The results of clustering twenty-one (28) computers shown in Figure 7 are entirely different from the other type of fixed assets. All of the computers in the ICT Equipment using the ECIk and KMA algorithms resulted to similar repair. Clustering the second time, the results of the KMA changed and the ECIk is still similar. The clustering results of the ECIk as compared with KMA for ICT Equipment are highly reasonable and more reliable because of consistency.



Fixed Asset ID	Name	Category	Subcategory	Usage Factor	Age Factor	Level of Maintenance in K...	Level of Maintenance in E...
PUP-1-1-3055	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	4.0	0.3	MAJOR	MAJOR
PUP-1-1-3056	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	4.0	0.3	MAJOR	MAJOR
PUP-1-1-3057	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	4.0	0.3	MAJOR	MAJOR
PUP-1-1-3058	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	5.0	0.3	MAJOR	MAJOR
PUP-1-1-3059	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	3.0	0.3	MAJOR	MAJOR
PUP-1-1-3060	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	3.0	0.3	MAJOR	MAJOR
PUP-1-1-3061	CPU HP COMPAQ 6200 PRO	ICT EQUIPMENT	COMPUTER	4.0	0.3	MAJOR	MAJOR
PUP-1-1-3306	LABORATORY DEKTOP CO...	ICT EQUIPMENT	COMPUTER	1.0	0.6	MAJOR	MAJOR
PUP-1-1-3307	LABORATORY DEKTOP CO...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3308	LABORATORY DEKTOP CO...	ICT EQUIPMENT	COMPUTER	8.0	0.6	MINOR	MINOR
PUP-1-1-3309	LABORATORY DEKTOP CO...	ICT EQUIPMENT	COMPUTER	7.0	0.6	MINOR	MINOR
PUP-1-1-3310	LABORATORY DEKTOP CO...	ICT EQUIPMENT	COMPUTER	7.0	0.6	MINOR	MINOR
PUP-1-1-3311	LABORATORY DEKTOP CO...	ICT EQUIPMENT	COMPUTER	5.0	0.6	MAJOR	MAJOR
PUP-1-1-3439	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3440	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3441	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3442	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3443	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3444	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-1-1-3445	COMPUTER DESKTOP INTE...	ICT EQUIPMENT	COMPUTER	6.0	0.6	MINOR	MINOR
PUP-4-1-1243	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	7.0	0.8	MINOR	MINOR
PUP-4-1-1244	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	7.0	0.8	MINOR	MINOR
PUP-4-1-1245	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	7.0	0.8	MINOR	MINOR
PUP-4-1-1246	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	8.0	0.8	MINOR	MINOR
PUP-4-1-1247	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	8.0	0.8	MINOR	MINOR
PUP-4-1-1248	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	8.0	0.8	MINOR	MINOR
PUP-4-1-1249	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	8.0	0.8	MINOR	MINOR
PUP-4-1-1250	KYOCERA ECOSYS LASER...	ICT EQUIPMENT	PRINTER	8.0	0.8	MINOR	MINOR

Figure 7. Screenshot of Clustering Result of ICT Equipment for Fixed Asset Monitoring System

Table 1 presented the twenty-eight (28) ICT equipment composed of the following: the first 7 are the CPU HP Compact, 7 Laboratory Desktop Computer-Dell and the next 7 Computer Desktop Intel under the Computer Sub category and the 7 Kyocera Ecosys Laser printer under the sub category Printer. When the assets under the ICT Equipment are simulated, the cluster id number 1 and 3, under ECIk are performing better in terms of speed as measured by run-time in millisecond but the cluster_id number 2 and 4 under KMA have fewer iterations. The simulation result of the ECIk and KMA varies throughout the batches of clustering the ICT equipment with respect to run-time and no. of iterations.

Table 1. Summary of the results with the run time and number of iterations for Information and Communication Technology (ICT) equipment.

Cluster_id	Algorithm	run_time (ms)	iteration	Number of ICT Equipment
1	KMEANS	2	2	28
2	ENHANCED	1	4	28
3	KMEANS	6	2	28
4	ENHANCED	1	4	28

Table 2. Average speed and iterations for Information and Communication Technology (ICT) equipment.

ALGORITHM	AVE SPEED	AVE ITERATION
KMEANS	4.00	2.00
ENHANCED	1.00	4.00

Table 3. Summary of the test result with run time and number of iterations for motor vehicle

Cluster_Id	Algorithm	Run-time (ms)	Iterations	Number of Motor Vehicle
9	KMEANS	1	3	7
10	ENHANCED	0	2	7
11	KMEANS	1	2	7
12	ENHANCED	1	2	7

Table 2 presented the average runt-time and iterations of the two algorithms. Results showed that the average speed is better for ECIk and average iteration is better performed by the KMA.

Table 3 presented the results of the clustering performance four (4) times for motor vehicle. The result of run-time for Cluster id no. 10 and the no. of iterations shows better performance for ECIk. Cluster_id no. 12 is equally fast with Cluster_id no. 9, 11 and 12 under the ECIk algorithm. With number of iterations, ECIk and KMA have similar performance.

Table 4 showed the average speed and iteration of the two algorithms based on the summary of the test results with run time and no. of iterations for motor vehicle. ECIk performance in terms of Average speed and iteration is better than KMA.

Table 5 determines if the clustering of certain equipment is repeated or duplicated as indicated by true. False indicates that the clustering of the equipment is not duplicated or repeated. The clustering of the forty-two (42) equipment for 42 times shows varying results for ECIk and KMA. From the clustering of id no. 1 to id no. 28, that represents the ICT equipment, the first eight (8) computers are clustered once with no duplicates at the 1st attempt for ECIk and but with duplicates with KMA. Clustering from id number 9 to 28 of ICT equipment obtained more favorable results for ECIk because of the no duplicates after the 2nd attempt unlike with the KMA results. Id number 30 to 32 of Office Equipment, 36 to 38 and 29, 33 and 34, 36 to 38, and 41 to 42 of Motor Vehicles clustered without duplicates for ECIk at the first attempt and the results for KMA vary. The clustering result in minor and major repair of twenty-eight (28) ICT equipment produces less duplicates for ECIk than KMA. And the clustering results of the seven Office Equipment and seven Motor Vehicles produced also less duplicates for ECIk. KMA constantly produces duplicates for all of the Equipment.

Table 6 showed the summary of consistency of the two algorithms. The result is based on the number of duplicates resulted from the clustering of 42 equipment for ECIk and KMA from Table 5. The table used the formula of the Consistency Percentage = ((Number of Assets – Count of Assets with Two Type cluster)/ (Number of Assets)) * 100, to determine how consistent in terms of duplicates is the two algorithms. Based from computation of the consistency percentage, the ECIk is more consistent than KMA. Once the result is obtained using ECIk, it is more stable than the result of KMA. KMA's result is less consistent because it generates more duplicates.

The clustering of all equipment using the Enhanced Initialization Centroids (ECIk) for K-means algorithm in terms of duplicates is far better than the traditional standard K-means algorithm. Duplicates referring to the number of times the same item is clustered into more than one type of maintenance level, major or minor repair at a given time until convergence is met. The result of ECIk made the fixed asset monitoring more consistent. Once, clustered using the ECIk,, the result of maintenance level to be adapted to the physical equipment is already fixed at a given time.

Table 4. Average Speed and Iterations for Motor Vehicle

ALGORITHM	AVE SPEED	AVE ITERATION
KMEANS	1.00	2.50
ENHANCED	0.50	2.00

Table 5. Detailed consistency of ECIk and KMA algorithm in terms of duplicates for all equipment.

KMEANS			ENHANCED		
ID	CLUSTER	DUPLICATES	ID	CLUSTER	DUPLICATES
1	1	TRUE	1	1	FALSE
1	2	FALSE	2	1	FALSE
2	1	TRUE	3	1	FALSE
2	2	FALSE	4	1	FALSE
3	1	TRUE	5	1	FALSE
3	2	FALSE	6	1	FALSE
4	1	TRUE	7	1	FALSE
4	2	FALSE	8	1	FALSE
5	1	TRUE	9	2	FALSE
5	2	FALSE	10	2	FALSE
6	1	TRUE	11	2	FALSE
6	2	FALSE	12	2	FALSE
7	1	TRUE	13	1	FALSE
7	2	FALSE	14	2	FALSE
8	2	TRUE	15	2	FALSE
8	1	FALSE	16	2	FALSE
9	2	FALSE	17	2	FALSE
10	2	FALSE	18	2	FALSE
11	2	FALSE	19	2	FALSE
12	2	FALSE	20	2	FALSE
13	2	TRUE	21	2	FALSE
13	1	FALSE	22	2	FALSE
14	2	FALSE	23	2	FALSE
15	2	FALSE	24	2	FALSE
16	2	FALSE	25	2	FALSE
17	2	FALSE	26	2	FALSE
18	2	FALSE	27	2	FALSE
19	2	FALSE	28	2	FALSE
20	2	FALSE	29	2	FALSE
21	2	FALSE	30	1	FALSE
22	2	FALSE	31	1	FALSE
23	2	FALSE	32	1	FALSE
24	2	FALSE	33	2	FALSE
25	2	FALSE	34	2	FALSE
26	2	FALSE	35	2	FALSE
27	2	FALSE	36	1	FALSE
28	2	FALSE	37	1	FALSE
29	2	FALSE	38	1	FALSE
30	2	TRUE	39	2	FALSE
30	1	FALSE	40	2	FALSE
31	2	TRUE	41	1	FALSE
31	1	FALSE	42	1	FALSE
32	2	TRUE			
32	1	FALSE			
33	2	FALSE			
34	2	FALSE			
35	2	FALSE			
36	1	FALSE			
37	1	FALSE			
38	1	FALSE			
39	2	FALSE			
40	2	FALSE			
41	1	FALSE			
42	1	FALSE			

Table 6. Summary of consistency of ECIk and KMA algorithms for all equipment.

ALGORITHM	NO. OF INCONSISTENCY	NUMBER OF CLUSTERED FIXED ASSET	PERCENTAGE OF CONSISTENCY
KMEANS	12	42	71.43%
ENHANCED	0	42	100.00%

4. CONCLUSIONS

A new method of initializing centroids for K-means algorithm is integrated in fixed asset monitoring application. The K-Means partitioning based clustering algorithm required to define the number of final cluster (k) beforehand. In this study, the clustering is limited only to two (conforming to the Section 24, GAM of COA), minor and major repair types of maintenance. The application of K-means in the fixed asset monitoring, using depreciation factors which are the age and usage, proves that the K-means algorithm is also effective in accounting procedure. After testing, the over-all processing time of the enhanced initialization of centroids improves as compared to the traditional KMA. In the simulation of twenty-eight (28) ICT equipment the ECIk is getting slower, but not with the seven(7) office equipment and 7 motor vehicle. In terms of consistency, for all of the equipment, ECIk is more efficient and reliable.

The study has successfully developed ECIk with comparator method to obtain the highest and lowest pair of values as guide for determining the initial centroids. The random selection of initial seeds of KMA which led to less reliable results was eliminated by ECIk. The results were conclusive that the ECIk algorithm proved to be more consistent than K-means algorithm. The study further state that the comparative analysis between the two algorithms yielded conclusive ratings in favor of ECIk than KMA in terms of speed and duplicates but the number of iterations for ECIk was compromising when the dataset was getting larger. And after testing the developed system using different constraints, ECIk in FAMS application was still able to produce non-conflicting clustering of fixed assets to minor or major repair faster and with higher percentage of consistency. The performance of ECIk could be adapted to other applications in which consistency in final result is needed.

5. RECOMMENDATIONS

Based on the findings of the study, the researcher offers the following recommendations: The integration of the comparator method in the ECIk is proven more effective in clustering non- overlapping results with higher percentage of consistency. But for some datasets with similar values of usage and age factors, the determination of initial centroids is a big challenge. The ECIk was challenged to find better initial seeds and this could be improved further in the future research. The over-all performance of ECIk is better as compared with KMA in terms of duplicates, enhancement did not eliminate the duplication and of the clustering of the same equipment as seen in the table of the Duplicates. The increasing number of the datasets resulted to the occurrence of the duplicates. Duplicates slow the speed of clustering and it has an effect on the number of iterations. This effect, could be the focus of the next research. The ECIk was not able to lessen the number of repetitions for all of the equipment clustered. Since, iterations is a part of the KMA process, the study was able to reduce it for majority of the equipment, but not to all. The future research must have a focus on how the iterations using Euclidean distance will be reduced in the entire clustering process. The ECIK needs more study and enhancements to contribute to the g`lobal optimization of k-means algorithm. ECIK makes the clustering converge faster with consistency reaching local optimization. But it is not yet the best solution model.

6. REFERENCES

Aggarwal, N., Aggarwal, K., & Gupta, K. (2012), Comparative Analysis of K-means and Enhanced K-means Clustering Algorithm for Data Mining. *International Journal of Scientific and Engineering Research,* 3 (3)

Commission on Audit (2015). *Accounting Policies, Guidelines and Procedures, and Illustrative Accounting Entries*. (COA Circular No. 2015 -007 Volume 1) Prescribing the Government Accounting Manual for Use of all National Government Agencies

Da Costa, J.C., Ortigueira, M.D., & Batista, A. G. (2013). *K-means Clustering for Sleep Spindles Classification*. International Journal of Information Technology and Computer Science *(IJITCS)*, 10 (3), 77 – 85.

Daljit Kaur, Kiran Jyoti (2013). *Enhancement in the Performance of K-means Algorithm.* International Journal of Computer Science and Communication Engineering, 2 (1) 29 - 32.

Dhanachandra, N., Manglem, K., & Chanu, Y. (2015). Image Segmentation using K – means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, Volume 54, 764-771, http://dx.doi.org/10.1016/j.procs. 2015.06.090

Fabregas, A., Gerardo, B., & Tanguilig III, B., (2017). Enhanced Initial Centroids for K-means Algorithm. *International Journal of Information Technology and Computer Science (IJITCS)*, 9(1), 26-33, DOI: 10.5815/ijitcs.2017.01.04

Haraty, R., Dimishkieh, M., & Masud, M. (2015). An Enhanced K-means Algorithm for Pattern Discovery in Health Care Data. *International Journal of Distributed Sensor Networks*, Article ID 615740, 11 pages http://dx.doi.org/10.1155/2015/615740

Hong, L., & Yu, X., (2009). Application Research of k-means Clustering Algorithm in Image Retrieval System. *Proceedings of the Second Symposium International Computer Science and Computational Technology, Huangshan* 274-277, doi: 10.5120/17846-8790

Improved Outcomes software. *Overview of K-means Clustering.* Retrieved from http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/K-means_ Clustering_Overview.htmImproved Outcomes software.K-means clustering, Retrieved from *onmyphd.com/?p=k-means.clustering* Algorithm and Examples

Kushwah, S. P. S., Rawat, K., & Gupta P (2012). *Analysis and Comparison of Efficient Techniques of Clustering Algorithms in Data Mining*. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 1(3), 109- 113. National Budget (2002). *Manual on Disposal of Government Property*. (NB Circular no.425)

Puri, P, & Sharma, I. (2015). Enhancement in K-means clustering to Analyze Software Architecture using Normalization. *International Journal of Scientific and Engineering Research (IJSER)*, 604-611.

Rauf, A., Sheeba, Mahfooz, S., Khusro, S., & Javed, H. (2012), Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity. *Middle-East Journal of Scientific Research*, 12 (7): 959-963, DOI: 10.5829/idosi.mejsr.2012.12.7.1845