International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 02 - Issue 11 || November 2017 || PP. 53-60

Improving The Density Cluster Performance Of K-mean Clustering By Using Log Function

Henderi Henderi[1, 3], Sri Hartati[2],Suharto Suharto[2]
[1]*Doctoral Program in Computer Science Dept. of Computer Science and Electronics, Department, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2)] *Departement of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[3)]*Departement of Information Technology, STMIK Raharja, Tangerang, Indonesia*

Abstract: K-mean is a popular algorithm and is often used in classifying data. K-means clustering is an algorithm to classify or to group the objects based on attributes into K number of group. In K-mean, data becomes a member of the k^{th} cluster if the Euclidean distance of the data to the center of the k^{th} cluster is the smallest compared with the distances to other cluster centers. Although it's commonly used, K-mean potentially causes misinformation because the resulting cluster may have low performance. This study developed a log function algorithm in Euclidean Distance in K-mean (LED K-mean) to reduce misinformation potential in the cluster resulted from K-mean by improving the performance of cluster related with cluster density. LED K-mean is a modification of K-mean algorithm which is performed by adding log function in determining Euclidean distance. Experiment on LED K-mean was performed on 1.744 data of key performance indicators of study programs with 3 testing scenarios. Then, the cluster performance produced by LED K-mean was evaluated using cluster density approach by measuring average values of entropy, center tendency, local density, and cluster diameter, and comparing them with the ones produced by K-mean. The result of 24 times of LED K-mean experiment in clustering the performances of key performance indicators in 3 study programs showed that the cluster performance significantly improved, as indicated by the average values of entropy, center tendency, local density, and cluster diameter which were relatively smaller than the ones produced by K-mean. They were 0,4; 1,5; 1,6; 2,3 compared with 2,5; 13,8; 15,7; 27,0, respectively. In this contribution, the improved clustering performance K-mean in terms of cluster density is shown by reducing values of entropy, center tendency, local density, and cluster diameter distance.
Keywords: K-mean, LED K-mean, cluster performance, cluster density

I. INTODUCTION

Standard K-mean (K-mean) is a popular algorithm and is often used in classifying data. K-means clustering is an algorithm to classify or to group the objects based on attributes into *k* number of group. K-mean is a partitioning method which uses average cluster values which are organized in such a way to be the same cluster to represent all data in the cluster.

Although it's commonly used, K-mean is inefficient for clustering complex data set and may cause error in classification [23]. Moreover, K-mean algorithm also has the limitation of potentially causing misinformation because the resulting cluster may have low performance.

This study discusses the development of log function algorithm in Euclidean Distance in K-mean (LED K-mean) to solve the potential weakness in K-mean by improving the performance of clustering result from the perspective of cluster density.

LED K-mean algorithm was developed in this study in line with the study of Stanica, et al. [5] which focuses on high cluster density. However, in their study, cluster performance still can be improved in terms of cluster density.

In the present study, cluster performance was represented by the average values of entropy, center tendency, local density, and cluster diameter. Cluster is said to have good performance if the values of entropy, center tendency, local density and diameter are smaller [21].

II. LITERATURE STUDY

There are some studies related with local density. Gaussian approach was performed by Mukherjee and Senguptra [1], Gaussian Mixture approach by Mukherjee and Datta [3], dan Mukherjee, et al. [10]. In this research group, cluster performance is evaluated using different approaches, including local spatial. Cluster performance in this research group still can be improved in terms of cluster density.

Studies on local density by non-linear variational approach was performed by Carrillo, et al. [2], heuristic approach by Chandrakala and Sekhar [4], segment-based approach by Haouari, et al. [7], and layerbased approach by Eilers, et al, [9]. This group has discussed the aspect of density, but the cluster performance in terms of cluster density also still can be improved.

Other sources, k-nearest neighbors approach was perdormed by Thang, et al. [8], vector quantization approach by Silva and Wunsch [6], Gaussian function approach by Marconcini, et al. [11], Bessel function approach by Wong, et al. [15], and Graphene plasmonics approach by Chen, et al [16]. The research group has limitation in cluster performance, especially large cluster.

On the other hand, density based studies by k-NN approach was performed by Yuan, et al. [12], K-mean approach by Yudong, et al, [18] and Kurass, et al. [20], and connectivity matrix approach by Tasdemir and Merenyi [19]. Some other studies on using log function are by Nigam, et al. [17], Liu, et al.[13], and Sen, et al. [14]. The research limitations in this group include not handling varying density.

Considering the literature review, it's necessary to study how to improve clustering performance in terms of cluster density by reducing the values of entropy, local density, center tendency, and cluster diameter.

III. ALGORITHM DEVELOPMENT

In K-Mean algorithm, the clustering process on data starts by identifying the data to be classified. In Equation (1), the data to be classified is expressed as x_{ij} (i=1,...,n; j=1,...,m) with n being total data to be classified and m being total variable.

At the start of iteration, the center of each data group was set randomly, c_{kj} (k=1,...,k; j=1,...,m), while the distance between every i^{th} data (x_i) to the center of the k^{th} cluster (C_k), was calculated using Euclidean Distance formula in Equation (1). The distance of the i^{th} data was calculated against every centroid cluster, and the i^{th} data will be a member of cluster k which had minimum distance to centroid cluster k.

1 Determine total cluster(k)
2 Set centers of clustersrandomly
3 Calculate distance of each data to cluster center: $d_{ik} = \sqrt{\sum_{j=1}^{m}(x_{ij} - c_{kj})^2}$
4 Classify data into clusterbased on the shortest distance
5 Calculate new centers of clusters
6 Repeat steps 3 to 5 until no data changed cluster

Figure 1. K-mean algorithm

Data became a member of the j^{th} cluster if the distance of the data to the center of the j^{th} cluster was the smallest compared with the distance to other cluster centers according to the results of calculation of Euclidean Distance by Equation (1).

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{kj})^2} \quad \text{........................ (1).}$$

In Equation (1), Euclidean Distanceis expressed as d_{ik} , data to be clustered is expressed as x_{ij} , and the cluster center is expressed as c_{kj} .

In Fig. 1, data classification is performed based on similarity, which is reflected by the distance of data to cluster center in the 3^{rd} and 4^{th} steps which potentially produces low performance, causing error in interpreting the resulting cluster. The potential problem can be reduced by developing a method which can improve the performance of the cluster produced by K-mean.

For this end, this study developed LED K-mean algorithm to improve the clustering performance in K-mean by modifying Euclidean Distance calculation by adding log function into Equation (1) to be Equation (2).

$$d_{ik} = log \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{kj})^2} \quad \text{.................. (2).}$$

Modification of Euclidean Distance model in K-mean algorithm into Equation (2) aimed to shorten the distance of each data to cluster center to improve the resulting cluster density.

In Fig.2, the basic difference between the LED K-mean developed by K-mean is in the third step which serves to make Euclidean Distance value relatively smaller. The step aimed to improve the cluster performance of the resulting cluster density. In the developed LED K-mean, the process of reducing the distance of data with the center of the cluster was performed in the beginning of iteration, and the result was the benchmark in making cluster in the first iteration.

1 Determine total cluster(K)
2 Set centers of clustersrandomly
3 Calculate distance of data to cluster center; $d_{ik} = log\sqrt{\sum_{j=1}^{m}(x_{ij} - c_{kj})^2}$
4 Classify data into clusterbased on the shortest distance
5 Recalculate new centers of clusters
6 Calculate distance of data to cluster center; $d_{ik} = \sqrt{\sum_{j=1}^{m}(x_{ij} - c_{kj})^2}$
7 Repeat 4th to 6th step until no data changed cluster

Figure 2. LED K-mean algorithm (proposed)

Next, the Euclidean Distance value produced in the first iteration became the benchmark in determine new cluster center point. The calculation of Euclidean Distance in the second iterations and so on used Euclidean Distance model on K-mean until stable cluster was produced.

3.1.Cluster Performance

Cluster performance was evaluated to determine the consistency of LED K-mean algorithm in improving cluster performance, which is indicated by reduced values of entropy, local density, center tendency, and cluster diameter it produced. Cluster performance is better if the values of entropy, local density, center tendency, and cluster diameter are smaller [21].

3.2.Entropy

Entropy is an external evaluation method to evaluate how good the clustering result which uses class data label as external information is. Cluster performance is better if the entropy values of data objects in each cluster are smaller. The value of entropy cluster was calculated using Equation (3) and average value of entropy was calculated using Equation (4).

$$e_i = -\sum_{j=1}^{L} p_{ij} \; log_2 p_{ij} \quad \text{.............................. (3).}$$

$$\underline{e}_i = \frac{(-Log_2 p_{ij}) + (-Log_2 p_{ij} \ldots n)}{\sum n} \quad \text{...................... (4).}$$

In Equation (3), the value of entropyof every clusteris e_i, where p_{ij} is calculated as $p_{ij} = m_{ij} / m_i$, with m_i as total data in cluster i, and m_{ij} is the result of summation of data in cluster i. In Equation (4), the average value of entropy clusteris $\underline{e}_i$, where n is the resulting cluster, and $\sum n$ is total cluster.

3.3. Local density

Local density is the average distance of objects in $N_k(o)$, *where* $N_k(o)$is the average density. Local density can be expressed as average value of data objects in cluster. The value of local density was calculated using Equation (5).

$$\underline{X} = \frac{\sum_{i=1}^{N} Xi}{N} \quad \text{.................................... (5).}$$

$$\underline{X}_i = \frac{\left(\frac{\sum_{i=1}^{N} Xi}{N}\right) + \left(\frac{\sum_{i=2}^{N} Xi}{N}\right) + \left(\frac{\sum_{i=n}^{N} Xi}{N}\right)}{\sum n} \quad \text{............(6).}$$

In Equation (5), local density cluster is expressed as $\underline{x}$, calculated as total values of data in cluster which is expressed as $\sum_{i=1}^{N} Xi$, where *N* is the amount of data in cluster. In Equation (6), the average value of local density is expressed as $\underline{X}_i$, where n is a cluster, and $\sum n$ is total cluster.

3.4. Centre tendency

Centre tendency is a data set for the most common and effective numerical measurement which is mean. However, center tendency cluster can be calculated by calculating mean, median and midrange of data objects in a cluster [21].

Median value of cluster was calculated using Equation (7) which is expressed as Med_c, where L_1 is the lower limit of median interval, N is total value in data set. $(\sum freq)e$ is total frequency of everything in the interval lower than median interval. $freq_{median}$ is frequency of median interval and width is width of median interval. Based on Equation (7) the average value of median is expressed in Equation (8), where c_k is the kth cluster, and $\sum k$ is total cluster.

$$Med_c = L_1 + \left(\frac{\frac{N}{2} - (\sum freq.)e}{freq_{\cdot median}}\right) width \text{ .... (7).}$$

$$\underline{Med}_c = \frac{(Med_{c1} + Med_{c2} + Med_{ck}}{\sum k} \text{ .... (8).}$$

In another section, midrange value is also used to measure center tendency for numerical data set. Midrange stated as Mid_c is the biggest and smallest average values in data set x_i, and was calculated using Equation (9). Next, the average value of Midrange is expressed as $\underline{Mid}_c$, and was calculated using Equation (10), where $\sum k$ is total cluster.

$$Mid_c = \frac{1}{2}[(max\ (x_i) + min\ (x_i)] \text{....... (9).}$$

$$\underline{Mid}_c = \frac{(Mid_{c1} + Mid_{c2} + Mid_{ck}}{\sum k} \text{................ (10).}$$

$$\underline{CT}_c = \frac{1}{3}(\underline{X}_i + \underline{Med}_c + \underline{Mid}_c) \text{.......... (11).}$$

Considering Equation (6), Equation (8) and Equation (10), the average value of center tendency is expressed as $\underline{CT}_c$, which is the average values of the resulting mean, median and midrange cluster. $\underline{CT}_c$ value was calculated using Equation (11).

3.5. Cluster diameter

Diameter value is max_{point} of data is a cluster. The cluster diameter value is expressed as max_{point} of data in all resulting clusters, and was stated with Equation (12), where x_i is all data in a cluster (c_k).

$$D_{ck} = \max(x_i, c_k) \text{............................ (12).}$$

$$\underline{D}_c = \frac{(D_{c1} + D_{c2} + D_{ck})}{\sum k} \text{ .... (13).}$$

The average cluster diameter is expressed as $\underline{D}_c$, which is cluster diameter divided by total formed cluster. $\underline{D}_c$ was calculated using Equation (13), where $\sum k$ is total cluster.

IV. EXPERIMENT RESULT

LED K-mean algorithm was demonstrated for clustering data of performance indicators of computer system, informatics engineering and information system study programs in STMIK Raharja. The data was numeric as shown in Table 1. Column × represents data of performance target and column y represents data of performance achieved by study program.

Table 1. Form of data set

No.	x	y
1.	74.3	78.3
2.	78.0	81.7
3.	82.0	91.1
…	…	…
…	…	…
…	71.0	92.7
…	92.0	89.6
…	77.1	83.7

Experiment was performed using 3 scenarios with 1.755 data. Scenario 1: was performed on 8 data sets with total data 301 (the 1st to 8th data sets), scenario 2: was performed on 8 data sets with total data 628 (9th to 16th data sets), and scenario 3: was performed on 8 data sets with total data 815 (17th to 24th data sets).

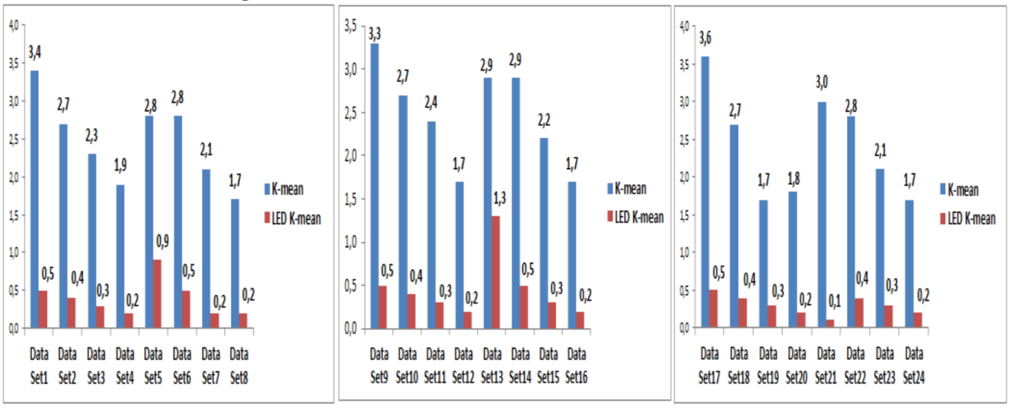
The experiment result of LED K-mean algorithm is presented in numerical reported adopted from the study of Marconcini and Macucci [11]. The experiment was performed to determine the consistency of LED K-mean algorithm which was developed to improve clustering performance.

Then, evaluation on cluster performance was performed by calculating average values of entropy, local density, center tendency, and cluster diameter produced by LED K-mean algorithm and comparing them with the ones produced by K-mean.

The consistency of LED K-mean algorithm in improving clustering performance was shown by consistency of average reduction of the resulting values of entropy, local density, center tendency, and cluster diameter compared with the ones produced by K-mean.

4.1. Entropy value of cluster

The entropy values of clusters from experiment of LED K-mean and K-mean algorithms in clustering data sets was calculated using Equation (3), while the average entropy values of clusterswere calculated using Equation (4). The average entropy value of clusters from experiment of LED K-mean and K-mean algorithms on every data set is shown in Fig.3.



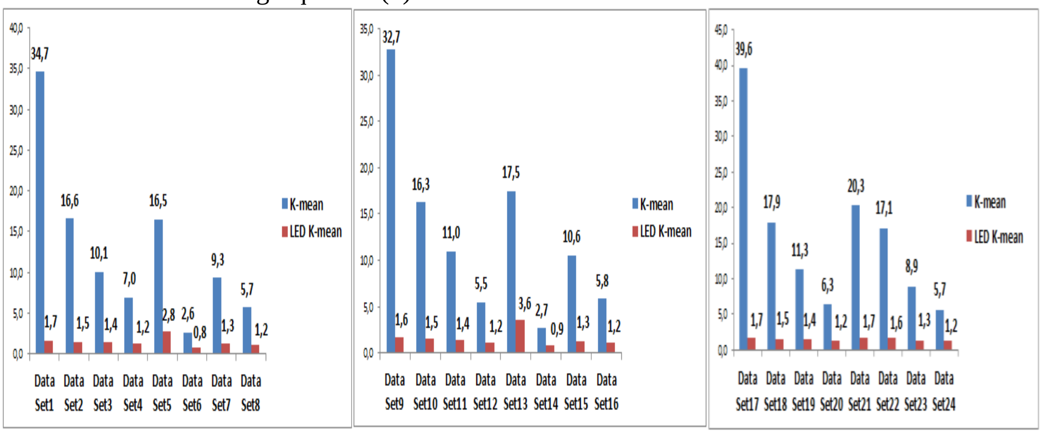
(a) (b) (c)

Figure 3. The average entropy values of the clusters

The experiment result revealed that the average entropy values of the clusters produced by LED K-mean were consistently smaller than then ones produced by K-mean. The average entropy values of the clusters produced by LED K-mean and K-mean in Fig. 3(a) used 301 data, Fig. 3(b) used 628 data, and Fig. 3(c) used 815 data, showing LED K-mean reduced average entropy values of the clusters produced by K-mean from 2,5; 2,5 and 2,4 to 0,4; 0,5 and 0,3, respectively. The experiment result in Fig. 3 shows that LED K-mean consistently reduced the entropy values of the clusters produced by K-mean. Reduced entropy value in the experiment seemed unaffected by total data in a data set.

4..2. Local density value of cluster

Local density values of cluster from experiment of LED K-mean were calculated using Equation (5). In another section, the average local density values of cluster from experiment of LED K-mean and K-mean algorithms were calculated using Equation (6).



(a) (b) (c)

Figure 4. The average local density values of the clusters

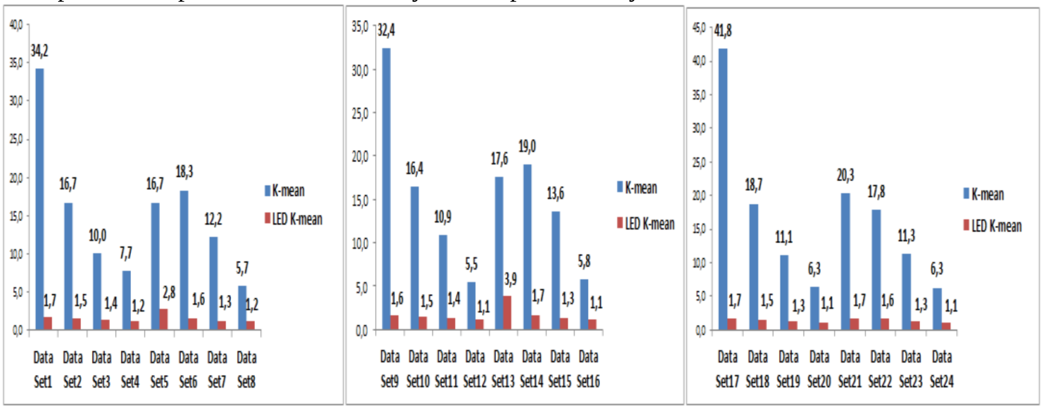
Fig. 4 shows that LED K-mean algorithm consistently reduced the average local density values of the clusters produced by K-mean. Consistency of reduced average local density values of the clusters is shown in Fig. 3(a) using 301 data, Fig. 3(b) using 628 data, and Fig. 3(c) using 815 data.

The comparison of average local density values of the clusters produced by LED K-mean with K-mean based on total experiment data was 12,8; 12,8 and 15,9 compared with 1,5; 1,6 and 1,5, respectively. The experiment result revealed that LED K-mean improved the performance of density cluster by reducing the average local density values of the clusters produced by K-mean by 89,1%.

4.3. Centre tendency value of cluster

The average center tendency values of the clusters from LED K-mean and K-mean algorithms were calculated by Equation (11). Fig.5 shows that LED K-mean consistently reduced the average center tendency values of the clusters produced by K-mean.

The average center tendency values of the clusters from LED K-mean and K-mean algorithms are presented in Fig. 3(a) using 301 data, Fig. 3(b) using 628 data, and Fig. 3(c) using 815 data. The experiment result showed that LED K-mean consistently improve cluster performance by reducing the average center tendency values by 89,9% from the ones produced by K-mean. It supported the evidence that LED K-mean algorithm improved the performance of density cluster produced by K-mean.

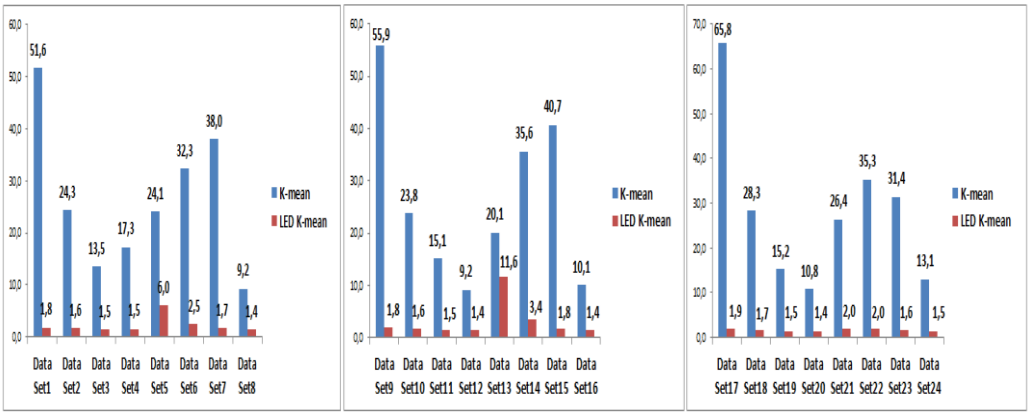


(a) (b) (c)

Figure 5. The average center tendency values of the clusters

4.4.Cluster diameter value

The average cluster diameter value was calculated by Equation (13). The experiment result in Fig. 6 showed that LED K-mean produced smaller average cluster diameters than the ones produced by K-mean.



(a) (b) (c)

Figure 6. The average cluster diameter values

Fig. 6 reveals that LED K-mean consistently improved the performance of density cluster by reducing cluster diameter values produced by K-mean. The comparison of the average cluster diameter values from K-mean algorithm with the results of LED K-mean in Fig. 3(a), Fig. 3(b) and Fig. 3(c) based on total data were 26,3; 26,3; and 28,3 compared with 2,3; 3,1; and 1,7, respectively.

The experiment result in Fig. 6 showed that LED K-mean consistently improve cluster performance produced by K-mean by reducing the average cluster diameter values.

V. DISCUSSION

The average values of entropy, center tendency, local density, and cluster diameter resulted from the experiment on 24 data sets with a total of 1.744 data using LED K-mean and K-mean with 3 scenarios are shown in Table 2.

The experiment result showed that LED K-mean algorithm in this study produced smaller average of entropy values than K-mean which was 0,4 to 2,5. It showed that LED K-mean algorithm improved cluster performance K-mean by 84,0% in terms of entropy. LED K-mean algorithm in this study is in line with ModEx algorithm developed by Rahman, et al [24] because they both improve cluster performance.

Table 2. The comparison of the average values of entropy, local density, center tendency and cluster diameter produced by K-mean and LED K-mean

	K-mean				LED K-mean			
Key Performance Indicator (Data Set)	**E***	**LD***	**CT***	**D***	**E**	**LD**	**CT**	**D**
Lecturer's Scientific Publication	3,4	35,7	36,1	57,8	0,5	1,7	1,7	1,8
Lecturer Quality Index	2,7	16,9	17,3	25,4	0,4	1,5	1,5	1,6
Lecturer's Discipline	2,1	10,8	10,7	14,6	0,3	1,4	1,4	1,5
Lecturer's Teaching Index	1,8	6,3	6,5	12,4	0,2	1,2	1,2	1,5
Final Assignment Supervising	2,9	18,1	18,2	23,5	0,8	2,7	2,8	6,5
Student's Discipline	2,8	7,5	18,4	34,4	0,5	1,1	1,6	2,6
Student's GPA	2,1	9,6	12,4	36,7	0,3	1,3	1,3	1,7
Graduates' GPA	1,7	5,7	5,9	10,8	0,2	1,2	1,2	1,4
Average	**2,5**	**13,8**	**15,7**	**27,0**	**0,4**	**1,5**	**1,6**	**2,3**

* E = Entropy, LD = Local Density, CT = Centre Tendency, D = Diameter

In terms of local density, the experiment result showed that the average local density value produced by LED K-mean was relatively smaller than the one produced by K-mean,which was 1,5 to 13,8. It showed that LED K-mean algorithm improved cluster performance K-mean by 89,1% in terms of local density.

In a different section, LED K-mean algorithm produced smaller average of center tendency value than the one produced by K-mean, which was 1,6 to 15,7. The experiment result revealed that LED K-mean algorithm improve cluster performance K-mean by 89,8% in terms of center tendency.

In terms of cluster diameter, LED K-mean algorithm produced smaller average cluster diameter, which as 2,3 to 27,0 produced by K-mean. It showed that LED K-mean improved cluster performance K-mean by 89,8% in terms of its diameter.

Overall, the experiment showed that the LED K-mean algorithm produced in this study improved the clustering performance of K-mean in terms of cluster density. This was shown by reduced average values of entropy, local density, center tendency, and cluster diameter produced by K-mean which were 2,5; 13,8; 15,7; 27,0 and became 0,4; 1,5; 1,6; 2,3, respectively.

VI. CONCLUSION

This study produced LED K-mean method which serves to improve clustering performance in terms of the resulting cluster density by adding log function to Euclidean Distance function in K-mean algorithm. In this contribution, the improved clustering performance by 88,6% compared K-mean in terms of cluster density is shown by reducing values of entropy, center tendency, local density, and cluster diameter distance.

Future research should be performed using more data types and greater amount of data to examine whether heterogeneous data affects the consistency of LED K-mean in improving cluster performance in terms of cluster density and in handling density cluster variation.

REFERENCES

[1]. Mukherjee, A., and Sengupta, A., Estimating the probability density function of a nonstationary non-gaussian noise, Transaction on Industrial Electronics, 57(4), 2010, 1429-1435.

[2]. Carrillo-Nu˜ nez, H., Magnus W., and Peeters, F.M., A non-linear variational principle for the self-consistent solution of poisson's equation and a transport equation in the local density approximation, International Conference on Simulation of Semiconductor Processes and Device, 2010, 171-174.

[3]. Mukherjee A., and Datta U., "Real time probability density function estimation in sensor networks", Sixt International Conference on Wireless Communication and Sensor Network, 2010, 1-4.

[4]. Chandrkala S. and Sekhar C.C., A density based methode for multivariate time series clustering in kernal feature space, IEEE International Joint Conference on Neural Networks (IEEE Word Congress on Computational Intellgence), 2008, 1885-1890.

[5]. Stanica R., Emmanuel Chaput, E., and Andr´ e-Luc Beylot, Local density estimation for contention window adaptation in vehicular networks, IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications, 2011, 730-734.

[6]. Leonardo Enzo Brito da Silva, E. L., and Wunsch II C. Donald, Multi-prototype local density-based hierarchical clustering, International Joint Conference on Neural Network (IJCNN), 2015, 1-9.

[7]. Noureddine Haouari N., Moussaoui, S., Guerroumi M., and Senouc Mohammed, S., Local density estimation for VANETs, Global Information Infrastructure and Networking Symposium (GIIS), 2016, 1-6.

[8]. Thang, V.V, Fantiukhin D.V., and Nazarov, A.N., FLDS: Fast outlier detection based on local density score, International Conference on Engineering and Telecommunication (EnT), 2016, 137-141.

[9]. Eilers S, Yi Du, and Dou Shi, X., Pb Thin Films on Si(111): Local density of states and defects, International Conferences on Nanoscience and Nanotechnology, 2014, 54-56.

[10]. Mukherjee, A., Mukherjee, D., Distributed probability density function estimation of environmental function from sensor network data, International Conference on Signal Processing, Image Processing, and Pattern Recognition (ICSIPRl), 2013, 346-350.

[11]. Marconcini P., and Macucci, M.,Effect of disorder on the local density of states of a graphene nanoribbon with a barrier, Proceedings of the 15th IEEE International Conference on Nanotechnology (IEEE NANO), 2015, 1126-1129.

[12]. Yuan, F., and Yang, L., Ge Yu, New density-based method for reducing the amount of training data in K-NN text classification, Proceedings of International Conference on Machine Learning and Cybernetics, 6, 2007, 3372-3376.

[13]. Kaiyi Liu, Zhiheng Zhou, Z., Zhuang, Y., Ming Dai, and Kuang, P., A log-function power-constrained algorithm for OLED displays, IEEE Region 10 Conference, 2015, 1-5.

[14]. Sen, N., T.K Basu, and Chakroborty, S., Importance of using log function to reduce the correlation between features in a multidimensional feature space for text-independent speaker identification, International Conference on Devices and Communications (ICDeCom), 2010, 1-5.

[15]. Wong Y.H., Penzin O., Braga N., and R.V. Mickevicius, Quanyum correction in AIGaN/GaN transistor simulations using modified local density approximayion (MLDA), IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC), 2018, 239-242.

[16]. Chen, Y.P., Wei E.I, Sha, Li Jun Jiang, and Jun Hu, Manipulating electromagnetic local density of states by graphene plasmonics, IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP), 2015, 1-3.

[17]. Nigam J., Dinesh Rai, and Tushir M., Application of log based kernel function in image segmentation, 3nd International Conferenc e on Computing for Sustaindable Global Development (INDIACom), 2016, 2615-2619.

[18]. Yudong G., Yanfang Li, Wang, Y., and Mingxin Liu,The application of an improved K-means clustering method in microarray gene expressing data, International Conferences on Biomedical Engineering and Computer Science, 2010, 1-4.

[19]. K. Tasdemir and E. Merenyi, Exploiting data topology in visualization and clustering of self-organizing maps, Transactions on Neural Networks, 20(4), 2009, 549-562.

[20]. Kurras M., Fähse S., and Thiele L., Density based user clustering for wireless massive connectivity enabling internet of things, IEEE Globecom Workshops (GC Wkshps), 2015, 1-5.

[21]. Han J., Kamber M., and Pei J.,Data mining: concept and techniques, Third Edition, Morgan Kaufmann Pubblishers, Elsevier, Waltham, USA, 2012.

[22]. Mirkin Boris, Clustering for data mining: Adata recovery approach", Taylor & Francis Group, LLC Chapman & Hall, London, UK, 2005.

[23]. Bo Wu and Wilamowski, B.M., A fast density and grid based clustering method for data with arbitrary shapes and noise", IEEE Transaction on Industrial Informatics, PP (99), 2016, 1

[24]. Rahman Md. A., Islam Md. Z., and Bossomaier, ModEx and seed-detective: two novel techniques for high quality clustering by using good intial seeds in K-means, Journal of King Saud University – Computer and Informatio Sciences, 27,2015, 113-128.