Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

¹ Implementation of K-Means Clustering Algorithm Using Java

² Prof. S.China Venkateswarlu¹, Prof. M.Arya Bhanu² and Prof.Yudhaveer Katta³

¹ JNTUH, HYDERABAD

Received: 3 August 2011 Accepted: 1 September 2011 Published: 11 September 2011

6 Abstract

3

4

Emergence of modern techniques for scientific data collection has resulted in large scale 7 accumulation of data pertaining to diverse fields. Conventional database querying methods 8 are inadequate to extract useful information from huge data analysis. Cluster analysis is one 9 of the major data analysis methods and k-means clustering algorithm Emergence of modern 10 techniques for scientific data collection has resulted in large scale accumulation of data 11 pertainting diverse felids. Conventional Data base methods are inadequate to extract useful 12 information from huge data banks. Cluster analysis is one of the major data analysis methods 13 and the k-means clustering algorithm is widely used for many practical applications. But the 14 original k-means algorithm is computationally expensive and the quality of the resulting 15 clusters heavily depends on the selection of initial cancroids. Several methods have been 16 proposed in the literature for improving the performance of the k-means clustering algorithm. 17 The k-means algorithm is computationally expensive and requires time proportional to the 18 product of the number of data items, number of clusters and the number of iterations. This 19 papert proposes a method for making the algorithm more effective and efficient. 20

21

22 Index terms— About four key words or phrases in alphabetical order, separated by commas.

23 1 INTRODUCTION

a) Module design and organization i. Fixed Transmission n this module the fixed transmission is computed by
retrieving each link which is limited to maximum signal ratio. Then the fixed routes are computed which give
the maximum transmitted power per node and is limited to hardware constraint. Then the distance between
the nodes is compute to calculate the end-to-end reliability. The end-to end reliability is a decreasing function,
which can be treated as the cost metric for route selection.

ii. End to End Reliability This module focuses on the problem of optimizing transmission power levels and route selection on an end-to end basis. This module minimizes the endto end power for a fixed route. The end-to-end route reliability under the optimal power allocation scheme is represented as any fixed route, different power allocation schemes result in different end-to-end reliability and power consumption. The next step is to retrieve the total bandwidth then the distance between nodes are minimized and reliability routes are computed.

34 iii. Outage Diversity

In the module the case of a point-to-point link, is considered and the trade-off between route outage and consumed power in a network setting. This type of analysis gives insight to how fast the end-to-end outage decreases as more power is spent on the transmission. First, we look at the case that the maximum transmitted power at each link is fixed. It is observed that the route selection does not have any effect on the form of this tradeoff. By selecting the optimal route, we minimized the end to end outage probability by minimizing.

This shows that as long as we limit our approach to a single transmitter and a single receiver per link, even under optimal power allocation and route selection, the trade-off maintains the same form as in the single link case. Phase 1 of the heuristic algorithm requires a time complexity of O(nkp) for finding the initial centroids, as the maximum time required here is for computing the distances between each data point and all other data-

 $_{44}$ points in the set D. In the original k-means algorithm, before the algorithm converges, the centroids are calculated

45 many times and the data points are assigned to their nearest centroids. Since complete redistribution of the data 46 points takes place according to the new centroids, this takes O(nkl), where n is the number of data-points, k is

⁴⁷ the number of clusters and l is the number of iterations. To obtain the initial clusters, algorithm 4 requires O(nk).

48 Here, some data points remain in its cluster while the others move to other clusters depending on their relative

49 distance from the new centroid and the old centroid. This requires O(1) if a data-point stays in its cluster, and

50 O(k) otherwise. As the algorithm converges, the number of data points moving away from their cluster decreases

s1 with each iteration. Assuming that half the data points move from their clusters, this requires O(nk/2). Hence

 $_{52}$ the total cost of this phase 2 of the heuristic algorithm is O(nk), not O(nkl). Thus the overall time complexity of

53 the heuristic algorithm becomes O(nkp).

In an Enhanced k-means approach ??4] we are some of the data points may become more closer to a different cluster and such points are redistributed accordingly. The entire process is repeated until no more data points cross cluster boundaries. The enhanced method is described as Algorithm 4. This method involves keeping track of the distance between each data point and the centroid of its nearest cluster.

58 During the subsequent iteration, instead of computing the distance of the data point from all cluster centroids, 59 its distance from the previous nearest cluster alone is determined. If that distance is less than or equal to the 60 previous nearest distance, the data point.

b) Algorithm: finding the initial centroids i. Finding Intial Centroids Input: $D=\{d1,d2,...dn\} // \text{ set of } n \text{ data}$ items K // number of desired cluster Output: A set of k initial centroids.

63 Steps:

1.Set m=1: 2. Compute the distance between each data point and all other data-points in the set D: 3. Find 64 the closest pair of data points from the set D and form a data-point set $Am(1 \le m \le k)$ which contains these 65 two data-points,. Delete these two data points from the set D; 4. Find the data point in D that is closest to 66 the data point set Am. Add it to Am and delete it from D: 5. Repeat step 4 until the number of data points 67 in Am reaches 0.75^* (n/k); 6. If m< k, then m=m+1, find another pair of data points from D between which 68 the distance is the shortest, from another data -point set Am and delete them from D, Go to step 4: 7. For 69 each data-point set Am $(1 \le m \le k)$ find the arithmetic mean of the vectors of data points in Am, these means 70 will be the initial centroids. Algorithm 3 describes the method for finding initial centroids of the clusters ??12]. 71 Initially, compute the distances between each data point and all other data points in the set of data points. Then 72 find out the closest pair of data points and form a set A1 consisting of these two data points, and delete them 73 74 from the data point set D. Then determine the data point which is closest to the set A1, add it to A1 and delete 75 it from D. Repeat this procedure until the number of elements in the set A1 reaches a threshold. At that point go back to the second step and form another data-point set A2. Repeat this till 'k' such sets of data points 76 are obtained. Finally the initial centroids are obtained by averaging all the vectors in each data-point set. The 77 Euclidean distance is used for determining the closeness of each data point to the cluster centroids. In the first 78 phase, the initial centroids are determined systematically so as to produce clusters with better accuracy ??12]. 79 The second phase makes use of a variant of the clustering method discussed in ??4]. It starts by forming the 80 initial clusters based on the relative distance of each data-point from the initial centroids. These clusters are 81 subsequently fine-tuned by using a heuristic approach, thereby improving the efficiency. The two phases of the 82 enhanced method are described below as Algorithm 3 and Algorithm 4. 83

The first step in Phase 2 is to determine the distance between each data-point and the initial centroids of 84 all the clusters. The data-points are then assigned to the clusters having the closest centroids. This results 85 in an initial grouping of the data-points. For each data-point, the cluster to which it is assigned (ClusterId) 86 and its distance from the centroid of the nearest cluster (Nearest_Dist) are noted. Inclusion of data-points in 87 various clusters may lead to a change in the values of the cluster centroids. For each cluster, the centroids are 88 recalculated by taking the mean of the values of its data-points. Up to this step, the procedure is almost similar 89 to the original k-means algorithm except that the initial centroids are computed systematically. The next stage 90 is an iterative process which makes use of a heuristic method to improve the efficiency. During the iteration, the 91 datapoints may get redistributed to different clusters. The method involves keeping track of the distance between 92 each data-point and the centroid of its presentnearest cluster. At the beginning of the iteration, the distance of 93 each data-point from the new centroid of its present nearest cluster is determined. If this distance is less than 94 or equal to the previous nearest distance, that is an indication that the data point stays in that cluster itself 95 and there is no need to compute its distance from other centroids. This results in the saving of time required 96 to compute the distances to k-1 cluster centroids. On the other hand, if the new centroid of the present nearest 97 cluster is more distant from the datapoint than its previous centroid, there is a chance for the data-point getting 98 included in another nearer cluster. In that case, it is required to determine the distance of the data-point from 99 all the cluster centroids. Identify the new nearest cluster and record the new value of the nearest distance. The 100 loop is repeated until no more data-points cross cluster boundaries, which indicates the convergence criterion. 101 The heuristic method described above results in significant reduction in the number of computations and thus 102 improves the efficiency. c) Algorithm: assigning data-points to clusters ??1] Assining DataPoints to Centroids 103 Input: manner.Once the initial centroids are thus determined, the distance between each data point and centroids 104 of all the clusters are determined and the data points are included in the nearest cluster. Cluster means are then 105 recalculated to find the new centroids. As a result of this, centroid. The algorithm proceeds in the following 106

max-imum time required here is for computing the distances be-tween each data point and all other data-points 108 in the set D. In the original k-means algorithm, before the algorithm converges the centroids are calculated many 109 times and thedata points are assigned to their nearest centroids. Since complete redistribution of the data points 110 takes place ac-cording to the new centroids, this takes O (nkl), where n is the number of data-points, k is the 111 number of clusters and l is the number of iterations. To obtain the initial clusters, Algorithm4 requires O(nk). 112 Here, some data points remain in its cluster while the others move to other clusters depending on their relative 113 distance from the new centroid and the old centroid. This requires O(1) if a data-point stays in its cluster, and 114 O(k) otherwise. As the algorithm converges, the number of data points moving away from their cluster decreases 115 with each iteration. Assuming that half the data points move from their clusters, this requires O (nk/2). Hence 116 the total cost of thisphase of the algorithm is O(nk), not O(nk). Thus the overall time complex2ity of the 117 enhanced algorithm (Algorithm 2) becomes O (n), since k is much less than n. 118 II. 119

120 2 CONCLUSION

In this section we have shown how testing is performed and different test cases are designed to test the system 121 for its performance as well as debugging process. The validation of the test cases is also shown. The k-means 122 algorithm is widely used for clustering large sets of data. But the standard algorithm do not always guarantee 123 good results as the accuracy of the final clusters depend on the selection of initial centroids. Moreover, the 124 computational complexity of the standard algorithm is objectionably high owing to the need to reassign the 125 data points a number of times, during every iteration of the loop. This Project presents an enhanced k-means 126 algorithm which combines a systematic method for finding initial centroids and efficient way for assigning data 127 128 points to clusters.

129 3 III. IMPLEMENTATION AND RESULTS

130 4 a) Introduction

In the module the case of a point-to-point link, is considered and the trade-off between route outage and consumed power in a network setting. This type of analysis gives insight to how fast the end-to-end outage decreases as more power is spent on the transmission. First, we look at the case that the maximum transmitted power at each link is fixed. It is observed that the route selection does not have any effect on the form of this tradeoff. By selecting the optimal route, we minimized the end to end outage probability by minimizing

¹³⁶ 5 IV. IMPLEMENTATION AND RESULTS

Algorithm 3 describes the method for finding initial centroids of the clusters ??12]. Initially, compute the 137 distances between each data point and all other data points in the set of data points. Then find out the closest 138 pair of data points and form a set A1 consisting of these two data points, and delete them from the data point set 139 D. Then determine the data point which is closest to the set A1, add it to A1 and delete it from D. Repeat this 140 procedure until the number of elements in the set A1 reaches a threshold. At that point go back to the second 141 step and form another data-point set A2. Repeat this till 'k' such sets of data points are obtained. Finally the 142 initial centroids are obtained by averaging all the vectors in each data-point set. The Euclidean distance is used 143 for determining the closeness of each data point to the cluster centroids. In the first phase, the initial centroids 144 are determined systematically so as to produce clusters with better accuracy ??12]. The second phase makes 145 use of a variant of the clustering method discussed in ??4]. It starts by forming the initial clusters based on 146 the relative distance of each data-point from the initial centroids. These clusters are subsequently fine-tuned by 147 using a heuristic approach, thereby improving the efficiency. The two phases of the VI. 148

149 6 CONCLUSION

In this section we have shown how testing is performed and different test cases are designed to test the system for its performance as well as debugging process. The validation of the test cases is also shown. The k-means algorithm is widely used for clustering large sets of data. But the standard algorithm do not always guarantee good results as the accuracy of the final clusters depend on the selection of initial centroids. Moreover, the computational complexity of the standard algorithm is objectionably high owing to the need to reassign the data points a number of times, during every iteration of the loop. This Project presents an enhanced k-means algorithm which combines a systematic method for finding initial centroids and ¹²

 $^{^{1}}$ © 2011 Global Journals Inc. (US)

 $^{^2 \}odot$ 2011 Global Journals Inc. (US) Global Journal of Computer Science and Technology Volume XI Issue XVII Version



Figure 1: ©

rootplacalhest	-																			F	8
96 of elements	97 in the	99 oluster	100 Lai 27	107																	
ements in th 3 50 of elements																					٩
ements in th 1 41 of elements				11 49																	
		SALK IND+																			
Al Ond.07 er On0.13 a On0.02 cot0localhos Newristic Re	tés tés rt -]# ti	Impeove	the effi	clency o	it the la	means clu	stering														
ver desired value of 8 sained clust 0 th clust M5333333333 S4000000000 75866686668 9866686668	t isti Her Gente Her Gente 1335 10007 16693																				
ements in th 1 23 45	e O th c 2 24 46	laster i 3 25 47	# 4 26 40	5 27 42	6 28 50	7 29 51	@ 30 52		10 32 54	11 33 55	12 34 56	13 35 57	14 36 50	15 37 59	16 18 60	17 39 61	10 40 62	19 41 60	10 43 64	11 43 65	3 4 4
45 67 09 111 133 of elements	68 90 112 134	69 91 113 135	70 92 114 136	73 93 118 137	73 94 116 138	73 95 117 139	74 96 118 140	75 97 119 141	74 90 120 142	77 99 121 145	78 100 123 144	79 3D1 323 345	80 102 114 146	81 103 128 147	83 104 124 148	83 105 127 149	84 106 128 150	85 107 129	64 85 103 130	87 109 131	
		DAR END.																			
al Ond.76 ec Gad.07 p On0.03 pot9localhos	12a 13 0																				
		182.168.2.1	80	otțicaliot	9 4	Ty ove-r	'wrt	- Y	3.5 - Perk		谢 untilled	- Faht	100	acto - Micros	dt Word				Addre	100	58 P

Figure 2:

Prostalecalho	st															6	0 18
																Land.	1
Nu wî elemen/	ta in the	e uliates															
		OFFAR DE															
real Dec.																	
ater 0m0.0																	
eye (bel.)																	
[root#localh																	
a Bearistic 1																	
Epter desire		clusters	k: 1														
the value of																	
Obtained clu																	
sew 0 th clu 5.30103092783		1920-13															
2.8865979391																	
4.9587626865																	
1.6958761886	\$97945																
nev i th clu 5.00544037735846 1.34037735846 1.5422441309 3.28847924525	58491 90567 43286	trotid 1.81															
elepents in t																	
51 1 74	52 75	53 76	54 77		56 79		59 61	190 182	67 114		54 86	65 101	60 90	69 91			
						104	105	104	108	109							
0, 142			145														
to of element	ta in the	t alwater	101 97														
elements in t	the 1 th	cluster															
																	2
1 19		15	26		28	29	30	31									
4 45 No of element	45	47	48				54										
C OI CIEDED	C# 312 CD		48- 22														
real Dad.	124-																
uner Out.																	
syz 0m0.4 [root#iocalhu																	
			127	100		-	Nor William									11 10 1	
start	California (stiĝierakes		(denn	A-Past									8.0 1	HE 101

Figure 3:

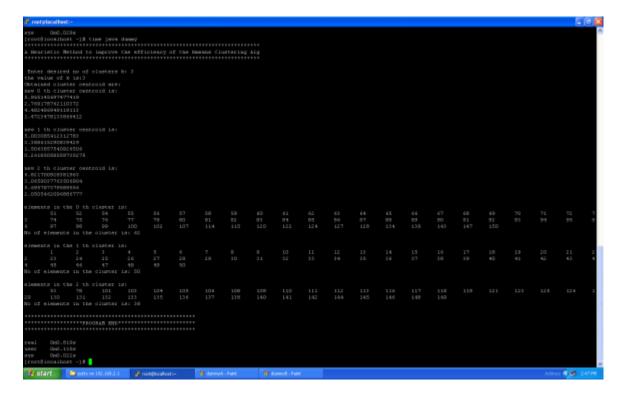


Figure 4:

rori fika a ho	1													
TOT HELE TO	8													
22531305190	1145154													
90762125781		1912 141												
75240103334														
4212453351														
4125536736														
a 1 th clu	iter cent	resd Las												
85146940711														
	Sec. 8													
esents in 1 51	ae 0 15 1	101	10:	124	\$15		108							
	112	213	116	117	118	119	121	123						
124	129	130	231	132	133	:35	116	117						
	141				346		145							
.cf element	11 15 the	citates												
ements in t														
							18							
40 of elment	11	46												
The Constantial	13 13	a started a												
esests in 1	he 1 th	cluster												
11	51	54												
	-94	-25	95		70		100	202						
116	125			124			174							
147 of element	150	-	101 53											
The Thread Ga	3 11 CZ	1110-111	121 14											
earets is t	the 3 th	eluster.												
of element	ia in the	0.48580												

al tad.:	12.54													
er (ad.)														
n 0x0.0														
	181]#													
(Protection)	-		14			1.0				1000				
start	Derte				20011	gi rattibi	calet.~		dek Refe					

Figure 5:

Section Section Section Section Sec	Ë ratil	bcelhni	e.															1	
att 04.104 04.015 <th>******</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>******</th> <th></th> <th>1</th>	******						******												1
Set 1.100 set 20:100 Set 1.100 set 20:100 Set 1.100 set 20:100 Set 20:100 set 20:1000 Set 20:100 set 20:1000 Set 20:1000 set 20:10000 Set 20:10000 Set 20:100000 Set 20:100000000000000000000000000000000000	******		imiga	HALE ON															
and a childs gene childs sectors sectors the means is the 16 so become to 5 sectors the means is the 16 so become to sectors set 15 solutes controls as: solutions set 15 solutes controls as: solutions solutions set 15 solutes controls as: solutions	******	*****	******		*******		******												
app: app: bit is App: bit is bit is	real	06.7	55																
Interflorence Section 1 # 110: provide the Minimum of the Newson Clatering Lig A Worker Ansatz A Section 2 # 10: Section 2 #	2241	Cm2.12	26e																
And and a state of a	<u>77</u> 2	DC.5																	
A functorial factorial conserves the efficiency of the bases (lattering Lig stress desired as of elements is 5 denoted control entropy of elements is 5 denoted control entropy of elements is 6 denoted contropy of elements is 6 denoted elements is 6 denoted elements is 6 deno																			
Amore Security 6s of Allowers 20:5 Store Security 6s Store Security																			
Cle mark 2 f E 12/3 Should Cluster (markund lan 5.1071280948075 5.10712804075 5.1071280400000000000000000000000000000000																			
247:247:25 1247:25 Contracts are: 247:047:257:0510 1.40:0712:0700010 1.40:0712:0700104 24.00:0477:050 24.00:0470:05001 25.00:050000001 24.00:0470:05000001 24.00:0470:05000001 24.00:0470:05000001 24.00:0470:05000001 24.00:0470:05000001 24.00:0470:05000001 24.00:0470:05000001 24.00:0410:04:04 24.00:0410:04:04 24.00:04:05000001 24.00:04:05000001 24.00:04:05000001 24.00:04:04:04:00:05001 24.00:04:04:04:00:05001 24.00:04:04:04:00:05001 24.00:04:04:00:00:0001 24.00:04:05000000 24.00:04:05000000 24.00:04:05000000 24.00:04:05000000 24.00:04:05000000 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:00:04:00:04:00 24.00:04:04:04:00:00:00 24.00:04:04:04:00:00:00 24.00:04:04:00:04:00:04:00 24.00:04:04:04:00:00:00 24.00:04:04:04:00:00:04:00 24.00:04:04:04:04:00:00:04:00 24.00:04:04:04:04:00:00:04:00 24.00:04:04:04:04:00:00:04:00 24.00:04:04:04:04:00:00:04:00 24.00:04:04:04:04:00:00:04:00 24.00:04:04:04:04:00:00:00:04 24.00:04:04:04:00:00:00:04 24.00:04:04:04:00:00:00:04 24.00:04:04:04:00:00:00:04 24.00:04:00:04:00:00:04:00 24.00:04:04:00:00:00:04 24.00:04:00:04:00:04:00:00:04:00 24.00:04:00:04:00:00:04:00 24.00:04:00:04:00:04:00:04:00:00:04:00 24.00:04:00:04:00:04:00:00:04:00 24.00:04:00:04:00:04:00:00:04:00:04:00:00:0	Ester	Sesire	t to of	elieters	12.5														
ear 5 to cluster control in: 5.10008000000000000000000000000000000000	the vel	at 12.)	£ 18:5																
6.171100000000 1.48710000000 1.487100000000 1.487100000000 1.4871000000000 1.487000000000 1.487000000000 1.10100000000000 1.1010000000000	Retaine	d clus	ter cest	trets are															
2.44010000000000000000000000000000000000				trid in															
4.9456.796779394 me 1 16 claste ventoud is: 5.070974504775 1.07098450775457 1.07098450775457 1.0709845075457 1.070984508504 me 2 16 claste ventoud is: 5.400001417128 1.409904000025 1.17099050000911 1.17099050000911 1.17099050000911 1.1709050000911 1.1709050000911 1.1709050000911 1.1709050000911 1.1709050000911 1.1709050000911 1.1499050000001 1.1709050000911 1.1499050000001 1.1499050000001 1.1499050000001 1.1499050000001 1.14990500000001 1.1499050000001 1.14990500000001 1.1499050000001 1.14990500000001 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.14990500000000 1.149905000000000000 1.1499050000000000000000000000000000000000																			
1.5771165903004 set 1 th clarte centruid is: 6.07364007751 0.073650077521 3.107580074601 set 2 th clarte centruid is: 6.400051417128 1.706970005015 1.70697000915 1.40677519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.44277519901601 3.442775199016001 3.44277519901601 3.4427751901601 3.44277519001601 3.44277519001601 3.44277519001601 3.44277519001601 3.4427751900000000000000000000000000000000000																			
new 1 th cluster centrons is: 0.02344:12441735 0.0234:12441735 0.0234:12401735 0.0234:12401735 0.0234:12401735 0.1035:00000015 0.1035:00000015 0.1035:00000015 0.1035:00000015 1.105:00000015																			
6.923442154912437 3.02345800196421753 3.1075590124617559 3.1075590124617559 3.107599012016125 3.10759901201255 3.10759901201255 3.1029901200090255 3.102900120009025 3.102900120009025 3.102900120009025 3.102900120009025 3.102900120009025 3.102900120009025 3.1029001200009025 3.1029001200009025 3.1029001200009025 3.1029001200009025 3.102900120000000000000000000000000000000	1.30703																		
6.923442154912437 3.02345800196421753 3.1075590124617559 3.1075590124617559 3.107599012016125 3.10759901201255 3.10759901201255 3.1029901200090255 3.102900120009025 3.102900120009025 3.102900120009025 3.102900120009025 3.102900120009025 3.102900120009025 3.1029001200009025 3.1029001200009025 3.1029001200009025 3.1029001200009025 3.102900120000000000000000000000000000000	ment t	b clust	er cest	read Las															
ELADERSMINAVAEZES 3.1071500020011 1a: 5.40000000011 1a: 5.40000000011 1a: 5.40000000011 1a: 4.708000000011 1a: 4.7080000000011 1a: 5.70100000000011 1a: 5.70100000000001 1a: 5.70100000000001 1a: 5.701000000000000000000 5.70100000000000000000000000000000000000																			
1.15715550001261 S.4000514017158 1.7687502010501 1.778575010500213 1.77857501050213 1.1259560000213 1.1259560000213 1.125957500002013 1.42427571595050 0.1251595750000 1.42425759159500 0.5788255751595 S. 4 Ta Claster centrol Ma: 	1.05199	669979	14557																
0er 2 to claste centrold La: 5.44000544(19188 5.44000544(19188 5.7440050430517 1.1709140900305057 0er 3 to claster centrold La: 6.799500009115 1.4245407427050009115 1.424540742705965 0.2013015745502407 0.401301545502407 0.2013015745502407 0.201574502407 0.201574502407 0.201574502407 0.201574502407 0.20157475024607 0.201574750407 0.201574750407 0.201574750407 0.201574750407 0.201574750407 0.20157475040000000000000000000000000000000																			
5.40005040197158 1.505000000000000 1.7050700000000000000 4.79507000000000000 1.1015000000000000 1.10150000000000 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.445477420059015 1.44545781644 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816416414 1.47564800816400814 1.47564800800800800800800800800800800800800800	2.10715	863784																	
1.5165554020675015 3.7487712216570079215 3.11755910009215 3.1429797129520079215 3.1429797129520079215 3.1429572129520079215 3.1429572129520079215 3.1429572129520079215 3.1429572129520079215 3.1429572129520079215 3.1429572129520079215 3.1429572129520079215 3.1429972129520079215 3.1429972129520079215 3.1429972129520079215 3.1429972129520079215 3.1429972129520079215 3.1429972129520079215 3.1429972129520079215 3.1429972105068725 4.145914022147 3.150 25 25 25 27 29 64 05 57 19 71 72 74 75 75 77 70 79 64 05 57 6 8 2 98 200 211 122 214 115 120 122 124 127 128 139 248 137 148 125 1 30 21 elements is the cluster is: 43 *Comments is the cluster is: 43 *Comments is the cluster is: 43 *Comments is the cluster is: 43				12015 18:															
8.7948712154510078 1.1259500009215 3.14279715020141 4.7959250009215 3.1427971502041 3.424547742105965 0.20131876532470 0.201318776532470 0.201318776532470 0.201318776532470 0.201318776552470 0.201318776552470 0.201318776552470 0.201318776552470 0.201318776552470 0.201318776552470 0.20131877770797707707970797070797070797079707079797070797970707979707079707970707979707079707970707979707079797070797079797070797970707979707079797070797970707979707079707079797070797970707979707079797070797970707970707970707979707079797070797970707970707970707979707079797070797970707979707079797070797970707970707979707079797070797970707970707970707979707079797070797970707979707079797070797970707970707979707079797070797970707979707079797070797970707970797970707979707079797070797970707970707979707079707079797070797970707970707970797070797070797070797070797070797070797070797070797070797070707070707079707070707070707070797070707970																			
1.110000000000000000000000000000000000																			
04F 3 Th Cluster Centrols 1s: 4.73902500009115 1.429551020912417 1.429551020921417 1.4295510205705502470 04F 4 Th Cluster Centrols 1s: 5.72116F390055005 3.5474829578416644 3.42004853051641644 3.42004853051641644 3.42004853051641644 3.42004853051641644 3.42004853051641644 3.42004853051641644 3.420098718700166470 elements in the 0 Th Cluster 1s: 5.21 S2 S2 S5 S7 S9 64 66 57 69 71 72 74 75 70 79 54 05 97 6 8 85 98 202 112 232 214 115 120 222 124 127 128 134 135 138 249 147 148 250 95 07 6 8 0 cf elements in the 1 sh element 1s: 108 100 204 208 004 508 120 120 110 213 114 117 218 119 123 123 124 126 124 120 313 112 112 122																			
4.72992300099213 3.142FT9T159012417 1.42FT9T159012417 5.0203308785392470 new 4 ta cluster destrout is: 5.721680290300000 8.54749297816544 5.721299787001066870 niements in the 0 ta cluster is: 5.22299787001066870 niements in the 0 ta cluster is: 5.25 52 52 55 57 19 64 56 57 67 72 74 75 75 77 70 75 64 05 97 6 8 82 98 901 513 522 124 115 520 122 124 127 128 124 535 239 243 147 148 250 8 01 elements in the 1 sh elumer in: 10 20 204 205 100 101 100 100 100 100 100 100 110 11	1.1.1255		50527																
8.44297971299224437 1.42450742.559555 0.2013018765592675 new 4 to cluster centrons Lat 5.201561876059060 3.479402277416544 3.450486205164164 5.2022997405064470 elements in the 0 to cluster 1s7 5.21 52 53 55 57 19 64 56 57 68 71 73 78 75 75 77 70 79 54 05 97 6 8 52 96 201 211 122 124 118 120 122 124 125 124 127 128 134 235 139 245 147 148 255 8 52 96 201 211 122 124 118 120 122 124 127 128 134 235 139 245 147 148 255 8 52 96 201 211 122 124 118 120 122 124 127 128 134 235 139 245 147 148 255 8 52 96 201 211 122 124 118 120 122 124 127 128 134 235 139 245 147 148 255 10 cf elements in the 1 th cluster 1s1 41 10 204 205 104 208 104 208 120 120 120 121 124 127 228 119 123 123 124 120 123 122 124 127 128				2015 18:															
1.41464074211550656 0.2013087765502670 3.5474828757416544 3.5772087701066870 alements in the 0 th cluster 1s: 1.42544825516516314 3.2172087701066870 elements in the 0 th cluster 1s: 1.42544127 128 119 123 123 124 125 125 125 125 125 125 125 126 129 120 123 125 125 125 125 125 125 125 125 125 125																			
0.2013018765392670 DeV 4 to cluster centrols is: 0.7015802605000 0.547402979141644 1.43048505164344 0.2020000000000000000000000 elements in the 0 to cluster is: 10 52 52 55 57 59 64 56 57 69 71 72 74 75 76 77 70 79 54 65 87 6 51 52 96 500 113 132 114 115 120 114 135 129 124 137 128 124 137 128 124 137 148 250 10 cf elements in the 0 luster is: 41 tlements in the 1 th elumer in: 10 10 104 106 106 108 109 110 111 114 117 128 119 123 123 124 129 128 129 130 131 132 1																			
24# 4 13 cluster centrois is: 5.7121651505005 8.74742957141644 5.7212299747051664733 mlements in the 0 to cluster is: 51 52 51 55 57 59 64 65 57 67 72 70 75 64 65 97 67 71 73 78 75 77 70 75 64 05 97 6 8 52 98 501 113 102 114 115 120 124 115 120 122 124 127 128 124 127 128 124 135 139 240 147 148 255 8 50 of elements in the laborer is: 10 10 104 105 105 105 101 110 110 111 114 117 128 119 123 123 124 120 328 182 132 1																			
5.727158270015960 3.547422975416544 3.4374229754165464 3.2122007310566437 elements in the 0 to cluster 1s: 51 52 52 55 53 59 64 56 57 95 71 73 74 75 75 77 70 75 54 05 97 6 8 52 96 206 111 212 114 115 110 112 114 117 128 114 115 118 148 250 15 8 02 elements 1x the cluster 1s: 41 10ements in the 1 so element 1s: 101 103 204 206 104 208 109 110 113 114 117 218 119 123 123 124 126 124 120 313 182 112 2	010032																		
5.727158270015960 3.547422975416544 3.4374229754165464 3.2122007310566437 elements in the 0 to cluster 1s: 51 52 52 55 53 59 64 56 57 95 71 73 74 75 75 77 70 75 54 05 97 6 8 52 96 206 111 212 114 115 110 112 114 117 128 114 115 118 148 250 15 8 02 elements 1x the cluster 1s: 41 10ements in the 1 so element 1s: 101 103 204 206 104 208 109 110 113 114 117 218 119 123 123 124 126 124 120 313 182 112 2	187 6 T	a class	ter cest	10010 L#1															
1.4384485551643434 3.212799737001664373 elements in 124 0 to cluster 1s: 21 52 52 55 57 59 64 56 57 69 71 72 74 75 76 77 70 79 54 05 87 6 8 82 96 501 113 132 124 113 120 124 123 124 127 128 124 535 139 140 147 148 250 80 cf elements in the cluster 1s: 41 tlements in the 1 th element is: 303 100 104 106 106 109 109 110 113 114 117 128 119 123 123 126 124 120 328 182 312 1																			
5.212299934051666933 mlements in the 0 to cluster is: 51 52 52 55 57 59 64 65 57 69 71 73 76 75 77 70 79 64 65 97 6 8 85 98 501 113 132 114 115 120 124 117 128 124 137 128 124 135 139 243 147 148 255 80 cf elements in the cluster is: 10 fements in the 1 th cluster is: 30 100 104 106 104 208 109 110 113 114 117 218 119 123 123 124 126 124 120 333 182 312 4	1,6749	957541	6644																
elements in the 0 th cluster is: 51 52 55 55 57 19 64 56 57 65 71 73 74 75 75 77 70 79 54 05 97 6 8 52 98 201 211 212 214 215 120 222 124 127 228 134 235 138 245 247 148 255 No of elements in the 1 sh element is: 41 nlements in the 1 sh element is: 303 103 204 205 104 208 103 110 213 114 117 218 119 223 123 124 129 323 182 312 32																			
51 52 53 55 57 19 64 56 57 69 71 73 76 75 77 70 79 54 65 97 6 8 96 301 111 1115 110 112 114 117 118 114 515 119 141 148 253 14 157 118 114 515 119 143 148 253 14 157 118 114 515 119 143 148 155 148 148 155 148 145 148 155 148 148 147 148 149 147 148	3.27279		16(473																
8 82 98 200 115 152 124 115 150 122 124 127 128 124 127 128 124 127 128 124 125 129 145 145 145 145 155 No of elements in the cluster is: 41 clements in the 1 th element is: 203 103 204 206 104 208 209 109 110 113 114 117 228 119 123 123 124 129 328 187 312 2	elesert																		
No of elements in the cluster is: 41 "Lements in the 1 sh cluster is: 203 103 204 206 104 208 209 209 120 213 214 217 228 119 223 123 326 224 129 333 282 332 2																			8
nlements in the 1 th clutter in: 303 103 304 108 106 308 109 100 110 310 310 110 110 110 110 110 110	Wo of .						1.1	115		941	** 1		114	112		948			
303 1C0 394 108 106 308 109 130 310 110 117 328 119 323 123 326 129 310 182 310																			
	e.earen																		1
	400	1	Dark		6	-	X471	ant	anglast-	 del-Net		Sec.	Part				436		SCR

Figure 6:

🖗 restálszaliest	*																		1	6 🗙
069 2 13 21081 5.44000524611 2.56058660189 3.78687822965 1.173599608080	1258 1903 10075 10527																			/
049 3 13 1114 4.79291501000 9.142975739930 1.414540749105 0.203311876525	2215 12417 19655	110 381																		
cen 4 t3 clost 5.22500234050 3.447499970410 2.490408285184 0.272790797050	2000 6664 63634	rold 3#1																		
elements in th	te 0 th i	cluster i																		
52																				8
5 92 No of elements	95 in the	102 rluster	111 131 61												147					
elements in th	te 1 th i	iluster i																		
																				1
33 236 No of elements	157 s in the	175 cluster	180 181 32	342																
elemento in th	te 1 th (:luster i																		
																				9
5 96 No of elements	97 13 134	59 tlugter																		
elements in th	te 3 th i	:luster																		
																				4
8 50 No of elements	t in the	eluster																		
elements in th	te 4 th (cluster (
																				4
5 41 No of elements	44 is the	45 cluster	47. 181 27																	
**********		SFLE END		 ******																
-01 000100																				
<pre>cesi 0s1.07 ustr 0s0.13 tys 0s0.03 [root#locs.hot</pre>	162 163																			
Retart	fistart Davie Dutys (R.M.2)			g rath	and the second	1	g al-het			Mar. Ht.			Tos-tas Stos-the			instant Allen 📢 🕬				

Figure 7:

<pre>collect.01.0.001 g general: methodships (# 150kb) 400 general: functional functions (# 150kb) 400 general: functions functions (# 150kb) functions functions (# 150kb) functions functions (# 150kb) functions (# 15</pre>	🖌 matrixed met -	
Ari Jopi An Wy i di Shuk 110 from 105 383.3.35 (mod bouthout - 1 from your Aug Arismus Robot to sprave the difference of the beam Clastering Hig Arismus Robot to sprave the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the beam Clastering Hig Arismus Robot to improve the difference of the Robot to improve the difference of the Robot to improve the difference of the Robot to improve to improve the Robot to improve to improve the Robot to	logia es: root	A 100
<pre>[nddboundf imp year way adapt if Arran Robot to segme to address of the beam functions if array of a rot events of</pre>	coct/192.105.2.100's password:	
A server which is agree to efficiency of the head linereig Hg Here Artise is of a last will of 1 and yet of 1 a		
Entry derived as of clusters h: 6 her wir of k init er (wollde coldection of a prove the efficiency of the besse flucturing hig the main of k init entry derived to grave the efficiency of the besse flucturing hig the main of k init ware of k inits control to grave the efficiency of the besse flucturing hig the state the board of the term is in the main of k init entry derived to distance is in the main of k inits entry derived to distance is in the state the board of the term is in the state term is in the state term is in the	<pre>tecorprocessor -it come leave draws.</pre>	
Enter detired as of clusters h: 6 he wise of h hand YCC The solution of the solution of the solution of the solution of the hand the solution of the soluti	a Secriptic Broad to improve the efficiency of the imenan Clustering alg	
<pre>An view of k usi Vi test 0v1100 fest 0v1100 fest</pre>		
<pre>An view of k usi Vi test 0v1100 fest 0v1100 fest</pre>		
<pre>Vet art 00/100 pre 00.0144 (pre)Active functions Active Active functions Active Activ</pre>		
rei m of 1.02 ref m 00.024 ref m 00.024 r		
<pre>xmm b0.100 typ b0</pre>		
Inveligencial of a run y we down Inveligence of a full full and run we down Inverse of the latter control are: are 0 the latter control are: are 0 the latter control are: Inverse of the latter of t		
<pre>1 Marinatio Mathia in improve the efficiency of the bases flucturing Hip Inter devices of x inc; Mathia flucture continuit are: Advanted continuit are: Advanted flucture continuit are</pre>	sys 0x0.014s	
<pre>1 Macintis Bethod to improve the efficiency of the heman flustening Hip Anticol A of clusters as 1 Anticol A of clusters as 1 Anticol A of clusters as 1 Anticol A of clusters control are: are 0 th cluster control are Anticol A of clusters Anticol A of cluster</pre>	[rootBlocalloat -]# clae jave dumey	
Exter desired to of clusters tr 1 Exter desinter desired to of cluster clusters tr 1 <td></td> <td></td>		
An vale of x 225 Analysis cluster central art: 2.54048452057583 2.54048452057583 2.54048452057583 2.5405855057583 2.5005700507763 2.5005850595058 2.5005850597583 2.5005850597583 2.5005710585759 2.5005710585759 2.5005710585759 2.5005710585759 2.50058508550477583 2.50058508550477583 2.50058508550477583 2.5005850850945297783 2.500585095519413 2.50058509519445 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.5005850951944 2.5005850951944	A Berclatic Bethod to implove the efficiency of the Roesne Clustering Fig	
An vale of x 225 Analysis cluster central art: 2.54048452057583 2.54048452057583 2.54048452057583 2.5405855057583 2.5005700507763 2.5005850595058 2.5005850597583 2.5005850597583 2.5005710585759 2.5005710585759 2.5005710585759 2.5005710585759 2.50058508550477583 2.50058508550477583 2.50058508550477583 2.5005850850945297783 2.500585095519413 2.50058509519445 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.50058509519443 2.5005850951944 2.5005850951944		
Resided -Luster centroid art: eW 0 % Schatter centroid art: Schatr	Exter desided to of clusters #: 0	
ee G Ab cluster centruid for L7122404074020 1.20071020074021 ee G Ab cluster centruid for E.20071020074021 ee G Ab cluster centruid for E.200710200740 1.20072000740 1.200740007400740 0.20074007750 4.20071040750 0.2007010040750 0.2007010040750 0.2007010040750 0.200701040750 0.200701040750 0.200701040750 0.200701040750 0.200701040750 0.200701040750 0.200701040750 0.200701040750 0.200701040750 0.20070100000000000000000000000000000000	the value of R 1st5	
E.17199440127480 .44144442404555 1.10407102007401 ass 3 th Slutes centrold int F.1701219510989 3.1270100407490 0.274409507504 0.27440950750504 0.27440950755541 1.444075514459775 0.28409481044775 0.28409481044775 0.28409481044775 0.28409481044775 0.28409481044775 0.28409481044775 0.28409481044775 0.2840948104475 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.2850948104404 0.285094810405 0.285094810405 0.285094810406 0.285094810406 0.285094810406 0.285094810406 0.285094810406 0.285094810406 0.285094810406 0.285094810406 0.285094840605 0.28509484060 0.285094840605 0.285094840005 0.28509484000000000000000000000000000000000		
2.44146462696555 1.2007022677621 2007022677621 201702219520989 5.21702219520989 5.21702219520989 5.21702219520989 5.2170220029578 5.21702019520929 5.217420029209585 5.217420920209595 5.217420920209595 5.217420920209595 5.217420920209595 5.217420920209595 5.217420920209595 5.217420920209595 5.2174209209595 5.2174209209505 5.2174209209505 5.2174209209505 5.2174209209505 5.2174209209505 5.2174209209505 5.2174209209505 5.2174209209505 5.217420050540100 5.217420050540100 5.217420050540100 5.217420050540100 5.21742050555 5.21742050555 5.21742050555 5.2174205555555 5.217420555555555555555555555555555555555555		
1.2157526484655 1.30577520477421 xee 3 % clustes centroid in: 5.1700129021999 5.51700129021999 5.5170012902100 0.2754279512025075 xee 2 % clustes centroid in: 6.4007394647519 0.26500000510477255 xee 3 % clustes centroid in: 6.4007394647519 0.26500000510477255 xee 3 % clustes centroid in: 6.4007394647519 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.26568483786107 0.26568483786107 0.26568483786107 0.26568483786107 0.26568483786107 0.26568483786107 0.26568483786107 0.26568483786107 0.265684837861007 0.265684837861007 0.265684837861007 0.265684837861007 0.26568483786100 0.26568483861007 0.26568483861007 0.26568483786100 0.26568483861007 0.26568483861007 0.26568483861007 0.26568483861007 0.26568483861007 0.26568483861007 0.265684838610000 0.265684838610000 0.2656848386100000 0.2656848386100000000000000000000000000000000000		
1.19971222971421 1.19971222971421 1.19071222971421 1.190712229715 1.19071229715715 4.421292212254 1.224512244502775 1.24451297151 1.244512971511 1.244512971511 1.24512244502775 1.2451244502775 1.25714297121255 1.350145071294144 1.55014507194054		
F.12T012219521959 F.12T012219521959 S.120006860795 S.120058670653508(D.2749409220939975 S.428294294258275 S.42829421582175 S.428297821582 S.42807522631781 S.42897522631781 S.42897522631781 S.428975226317411 S.428977449739112848 S.4289774	1.504071026077621	
F.12T012219521959 F.12T012219521959 S.120006860795 S.120058670653508(D.2749409220939975 S.428294294258275 S.42829421582175 S.428297821582 S.42807522631781 S.42897522631781 S.42897522631781 S.428975226317411 S.428977449739112848 S.4289774		
8.5127000565057963 9.274542002100390715 EME 7.3.210000650579 8.4258294218579 9.22452793 9.22452793 9.225000685194777938 EME 3.5.210006 Centrold 381 4.4607322248134120778 9.205685013964344 9.205685013954344 9.20568501395434 9.20568501395434 9.205586561350778 9.2055865613507704	ces 1 th cluster centricd is:	
1.4047542905335084 0.27494290539590710 X## 2 15 :20x8ter centroid is: 4.405942155374 1.400047318047519 0.2005084519497753 X#9 2 15 :21x8ter centroid is: 4.4057522441741 1.00145917838007 1.100145001204184 0.205584135220778 X#000784694202 5.4500078469412858 1.1005165701304		
0.274945005300390355 A4926983155276 1.2426292465325776 1.242637246532776 1.20500488594797753 See 3 th Dister Centroid and 4.40752204317411 1.505465931992807 1.50546502294114 5.1056854132672774 See 4 th Dister Centroid and 4.400794404242 1.576457197113555 1.55567258760356 1.195546478277004		
xew 2 %3 cluster centroid ss: 4400598155276 1.2461736453576 1.246173645359 0.200009865104977253 2.000009865104977253 2.000099865104977253 2.0000998650025944184 1.001460025984484 1.001660025944184 2.0058684336520774 2.0058684336520774 2.00594404242 1.0015667197113658 3.05566718270706		
2.4525963155276 1.424637246508775 1.425637535 0.00050085196977258 0.0005085196977258 0.46575522643741 1.0254654591783867 1.0254654591783867 0.2056555132672776 0.20565551326720776 0.20565551326720776 0.20565551326720776 0.2056551326720786 0.205657201563 1.1255657207506 0.205765132777006		
1.4245254463575 1.45047338647839 0.05090485194977253 Sev 3 th cluster centroid set 4.4673522493741 1.50145002294144 1.50145002294144 1.0005854135620774 Sev 4 th cluster denoratio set 4.460079440442 1.578045719713585 3.550567258740388 1.1955464762707004	new 2 th cluster centroid is:	
1.49047)18841819 0.2000081098119477201 04# 3 th cluster centroid ini 1.46875224811411 1.051449917893807 1.190165034126120714 0.1085834126120714 0.1085834126120714 0.1085834126120714 1.4700794404242 1.9704457197113585 1.9704457197113585		
D. 08190485194979783 A 48975204817411 A. 48975204817411 A. 100546907984087 A. 10058884336620774 Sew 4 th Dilate Cebrood is: S. 4620078406402 A. 2790487197113688 J. 5596719671006		
Mer 3 t3 cluster centróid so: 4.4675224917401 0.0554691780807 1.50546070240144 0.2056854132672078 Sem 4 t5 cluster Centroid So: 5.46000784806402 3.855967156760588 3.855967156760588		
A. 489751204311411 A.05764591788887 J.902145002994194 O.2085654326720776 SeW 4 th Dilates Cebsson iso S.480007845082402 J.5780457197123655 J.550647197123655 J.5506725477004	2.CO.20190121.01111020	
1.05164891785887 1.0501650022984184 0.0088584326500774 SeW 4 15 Cluster Geborood SBC 1.6780457197123658 3.8509672583740588 1.195548478277004	cev 2 th cluster centroid sp:	
1.96144002994184 5.009888433672774 5.46200784062402 1.9784457197113655 9.8559671543760586 1.1955484782707004	4.689752224327421	
0.0068684338630774 sew 4 th bluster Gebstood SBC 8.460007848063402 3.8559671568760588 1.1955484181707004		
xew 4 t3 blustep cebrood is: 5.440007848089401 1.5786457197128555 3.8509671568760158 1.199548(78270704)		
5.480007844063402 2.5780457197123685 3.5539072548740588 3.1935484782707004	0.00000120010014	
5.480007844063402 2.5780457197123685 3.5539072548740588 3.1935484782707004	sew 6 th cluster centrois ist	
3.850671568760158 1.1056484181707004	5.470007846078400	
1.1955484785707004	1.5780457197123668	
ff start Daub unit 2012 - Freedonier-	1-195566-16272-009	
ff state Denter all 1821 geneticator-		
	ff start potra 15: 16: 21 g northinator-	90 488

Figure 8:

Implementation of K-Means Clustering Algorithm Using Java c1,c2,.ck // set of k centroids Steps: 1. Compute the distance of each data point di $(1 \le i \le)$ to all the centroids of $(1 \le j \le k)$ as d(di, cj);2. For each data-point di, find the closest centroid cf and assign di to cluster j.

3. Set CllusterId[i]=j;

// j: Id of the clos- est

201 duster 4. Set Nearest _Dist[i] = d(di, cj);

Oct**5**beror each cluster j ($1 \le j \le k$) troids: 6. Repeat 7. For each data-point di. 8. Compute its distance is less than or equal to

64 the present nearest cluster:

9. If this distance is < or = to the present nearestdistance, the data-point satays in the cluster: Else 10. For every centroid cj $(1 \le j \le k)$ Compute the distance d(di, cj); Endfor; 11. Asign the data-point di to the cluster with the nearest centroid cj 12. Set Cluster ID[i]; 13. Set Nearest_Dist[i] = d(di, cf); 14. Endfor; 15. For each cluster j ($1{<}{=}$ j<=), recalculate the

centroids:

16. Until the convergence criterin is met.

Figure 9:

6 CONCLUSION

enhanced method are described below as Algorithm 3 and Algorithm 4.

The first step in Phase 2 is to determine the distance between each data-point and the initial centroids of 158 all the clusters. The data-points are then assigned to the clusters having the closest centroids. This results 159 in an initial grouping of the data-points. For each data-point, the cluster to which it is assigned (ClusterId) 160 and its distance from the centroid of the nearest cluster (Nearest_Dist) are noted. Inclusion of data-points in 161 various clusters may lead to a change in the values of the cluster centroids. For each cluster, the centroids are 162 recalculated by taking the mean of the values of its data-points. Up to this step, the procedure is almost similar 163 to the original k-means algorithm except that the initial centroids are computed systematically The next stage 164 is an iterative process which makes use of a heuristic method to improve the efficiency. During the iteration, the 165 datapoints may get redistributed to different clusters. 166

The method involves keeping track of the distance between each data-point and the centroid of its 167 presentnearest cluster. At the beginning of the iteration, the distance of each data-point from the new centroid 168 of its present nearest cluster is determined. If this distance is less than or equal to the previous nearest distance, 169 that is an indication that the data point stays in that cluster itself and there is no need to compute its distance 170 from other centroids. This results in the saving of time required to compute the distances to k-1 cluster centroids. 171 On the other hand, if the new centroid of the present nearest cluster is more distant from the data-point than its 172 173 previous centroid, there is a chance for the data-point getting included in another nearer cluster. In that case, 174 it is required to determine the distance of the data-point from all the cluster centroids. Identify the new nearest cluster and record the new value of the nearest distance. The loop is repeated until no more data-points cross 175 cluster boundaries, which indicates the convergence criterion. The heuristic method described above results in 176 significant reduction in the number of computations and thus improves the efficiency. 177

178 .1 V. OUTPUT

The modified algorithm is applied to multidimensional gene expression data taken from the UCI(university of california irvine) repository ??7]. The input dataare the iris data[10], the breast cancer data ??11], the e coli data[9], the echo cardiogram data[12], the yeast data[13] and the height-weight data obtained from the web site of disabled-world ??8]. The results are compared with that of the original k-means algorithm as well as Enhanced k-means algorithm. Tables 6.1 to 6.6 show the performance comparison of the three algorithms. Figures 6.1 to

6.6 illustrate that the modified algorithm provide better accuracy and efficiency compared to the k-means and enhanced k-means methods.