

clustering, each point belongs to a cluster. However, in soft clustering each point can belong to several clusters with different probabilities. As shown in Table 1, the EM is used for large data sizes unlike the K-Means family because the EM has been proposed as soft clustering and the K-Means is hard clustering in data mining. In regard to the analysis of existing parameters, the EM algorithm is better than other methods.

TABLE 1: Comparison of Clustering Algorithms

Algorithms	Accuracy	Convergence speed	data set size
K-Means	Low	Very low	Small
EM	High	High	Large
Elbow K-Means	Low	Low	Small
DBIK-Means	Moderate	Moderate	Small
Silhouette K-Means	Moderate	Moderate	Small
CHK-Means	Low	Moderate	Small

9. Conclusion

In this paper, the performance of the EM and the K-Means family clustering algorithms was compared according to parameters such as accuracy, convergence speed and data set size. The results indicated that the EM algorithm performs better compared to the K-Means family.

10. References

- [1] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proc. fifth Berkeley Symp. Math. Stat. Probab.*, vol. 1, no. 14, pp. 281–297, 1967.
- [2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [3] C. E. Rasmussen, "The infinite Gaussian mixture model," *Adv. Neural Inf. Process. Syst.*, pp. 554–560, 2000.
- [4] H. Shokrzadeh, S. Khorsandi, and A. T. Haghighat, "Optimized query-driven appointment routing based on Expectation- Maximization in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 35, no. 6, pp. 1749–1761, 2012.
- [5] K. Singh, D. Malik, and N. Sharma, "Evolving limitations in K-means algorithm in data mining and their removal," *IJCEM Int. J. Comput. Eng. Manag. ISSN*, vol. 12, no. April, pp. 2230–7893, 2011.
- [6] P. Bholowalia, and A.Kumar "EBK-Means : A Clustering Technique based on Elbow Method and K-Means in WSN," vol. 105, no. 9, pp. 17–24, 2014.
- [7] E. Rabiaa, B. Noura, and C. Adnene, "Improvements in LEACH based on K-means and Gauss algorithms," *Procedia Comput. Sci.*, vol. 73, no. Awict, pp. 460–467, 2015.
- [8] D. L. Davies. and D .W. Bouldin., "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no.2, pp. 224–227, 1979.
- [9] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 2321–7782, 2013.
- [10] P. J. Rousseeuw, "Silhouettes: A Graphic Aid to the Interpretation and Validation of Cluster Analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.
- [11] J. Baarsch and M. E. Celebi, "Investigation of Internal Validity Measures for K-Means Clustering," *Proceeding Internatinal MultiConference Eng. Comput. Sicientists*, vol. I, 2012.