ABSTRACT

Data Mining is the foremost vital space of analysis and is pragmatically utilized in totally different domains, It becomes a highly demanding field because huge amounts of data have been collected in various applications. The database can be clustered in more number of ways depending on the clustering algorithm used, parameter settings and other factors. Multiple clustering algorithms can be combined to get the final partitioning of data which provides better clustering results. In this paper, Ensemble hybrid KMeans and DBSCAN (HDKA) algorithm has been proposed to overcome the drawbacks of DBSCAN and KMeans clustering algorithms. The performance of the proposed algorithm improves the selection of centroid points through the centroid selection strategy. For experimental results we have used two dataset Colon and Leukemia from UCI machine learning repository.