IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, September 2024

A Novel Method for Large Datasets Mining by using Clustering Approach

Nitish Marathe[1], Dr. Harsh Lohiya[2], Dr. Rajendra Singh Kushwaha[3]
Research Scholar, Department of CSE[1]
Associate professor, Department of CSE[2]
Professor, Department of CSE[3]
Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India

Abstract: *Clustering is an unmanaged machine studying technique for discovering and grouping related information factors in massive datasets without regard for the end result. Clustering is wonderful in information mining because it enables the invention of companies and the identity of relevant distributions within the underlying facts. Historically used clustering techniques either choose spherical clusters with comparable sizes or are extraordinarily brittle in the presence of outliers. utilizes a mixture of random sampling and partitioning to manipulate massive databases. After partitioning a random sample of the statistics set, every partition is relatively clustered. After that, the partial clusters are clustered again to get the favored clusters. Numerous parallel methods based on the MapReduce structure were offered these days to deal with the scalability trouble due to growing data sizes. whilst huge records is clustered in parallel the usage of the K-Means set of rules, it's miles study again and again at some stage in each iterative step, drastically increasing each I/O and network fees. We provide a brand new series-based K-Means clustering method, dubbed CBKMeans, in this look at that successfully reduces statistics length even as improving clustering accuracy thru representative verification. Our experimental effects reveal that CBKMeans are extra efficient, scalable, and accurate than k-means.*

Keywords: Data mining, knowledge discovery, clustering algorithms, sampling

DOI: 10.48175/568