Abstract

Identifying clusters for huge datasets are useful for finding out attributes of a particular dataset and thereby providing insights for making effective decision making. In our previous work, we have proved the concept of clustering algorithms for huge datasets theoretically by applying small computations on the available datasets. In this paper, we extend the same work by applying Mathematical calculations for the datasets so as to prove the correctness of our previous work carried out. Our proposed method is applied to various datasets and proved K-Means algorithm mathematically and the experimental calculations performed on various clustering algorithms shows that our approach provides the new idea of clustering techniques that can be applied for any number of huge and complex datasets.

References


5. Robson L. F. Cordeiro et.al," Clustering Very Large Multi-dimensional Datasets with MapReduce." ACM- KDD’11, August 21–24, 2011, San Diego, California, USA.


7. Max Bodoia ," MapReduce Algorithms for k-means Clustering."


Index Terms

Computer Science Algorithms

Keywords

Machine Intelligence, Clustering Algorithms