Efficient and Scalable Approach for Large Datasets Using Extended Version of K-Mediod Algorithm

By

Swarndeep Saket Jayant Kumar
Enrollment No.: 140370702531

Guided by
Dr. Sharnil Pandya
Associate Professor
Department of Information Technology
Parul Institute of Engineering and Technology

A **Thesis** Submitted to
Gujarat Technological University
in Partial Fulfillment of the Requirements for
the Degree of Master of Engineering
in **Computer Engineering**

May – 2016

Department of Computer Science and Engineering
Parul Institute of Engineering & Technology
P.O: Limda, Ta.: Waghodia, Dist.: Vadodara

Efficient and Scalable Approach for Large Data Sets Using Extended Version of K-Medoid Algorithm

Submitted By

Swarndeep Saket

Supervised By

Dr. Sharnil Pandya

Associate Professor, Department of Information Technology

Parul Institute of Engineering and Technology,

Limda, Waghodia, Vadodara

ABSTRACT

Amongst various kinds of clustering techniques, partition based clustering are of immense importance and they are widely used in the modern days large data analysis. The K-Means and K-Mediod Algorithms in their existing format carry certain weakness. For example in case of K-Means algorithm 'deformation' and deviations may arise due to the misbehaviour and disruption in the computing process. Similarly in case of K-Mediod Algorithm a lot of iteration is required which consumes huge amount of time and their by reduces the efficiency of clustering. In addition, scalability is also not suitably adapted in the data analysis.

The present work seeks to modified and extend K-Mediod Algorithm to improve its efficiency in terms of reduction of execution time, so as to enable us to used it more effectively in case of Large datasets. The present study manifests that extended K-Mediod Algorithm are capable of handling large datasets with faster execution time, better scalability and overall improved efficiency.