## Cluster analysis using k-means and k-medoids

## B.Tech. Final Year Project Report

## BY

**ANISHA PAL**



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## NEOTIA INSTITUTE OF TECHNOLOGY, MANAGEMENT & SCIENCE

## JHINGA, AMIRA, D.H.ROAD, 24-PARAGANAS (S) - 743368, WB (INDIA)

## MAY, 2015

## cluster analysis using k-means and k-medoids

## A Major Project Report

### Submitted in partial fulfillment of the requirements for the award of the degree

*Of*

#### Bachelor of Technology

*In*

##### **COMPUTER SCIENCE AND ENGINEERING**

## BY

|  |  |  |
| --- | --- | --- |
| **ANISHA PAL** | **14400113005** | **Registration Number** |

**Under the Guidance of**

Prof. Subrata Bose



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## NEOTIA INSTITUTE OF TECHNOLOGY, MANAGEMENT & SCIENCE

## JHINGA, AMIRA, D.H.ROAD, 24-PARAGANAS (S) - 743368, WB (INDIA)

## November, 2016

# 



# CERTIFICATE OF ORIGINALITY

I hereby certify that the work which is being presented in the B.Tech. Final Year Project Report entitled **“Cluster analysis using K-means and K-medoids”,** in partial fulfillment of the requirements for the award of the **Bachelor of Technology in Computer Science & Engineering** and submitted to the Department of Computer Science & Engineering of Neotia Institute of Technology, Management & Science, West Bengal is an authentic record of my own work carried out from September, 2016 under the supervision of **Prof. Subrata Bose,** **CSE Department**.

The matter presented in this thesis has not been submitted by me for the award of any other degree elsewhere.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |
| --- |
| *Signature of Candidate* |
| **Anisha Pal** |
| **14400113005** |
| **Registration Number** |



**CERTIFICATE OF RECOMMENDATION**

This is to certify that the Project entitled “CLUSTER ANALYSIS USING K-MEANS AND K-MEDOIDS” has been submitted by **ANISHA PAL** under my guidance in partial fulfillment of the degree of Bachelor of Technology in Computer Science & Engineering of Neotia Institute of Technology, Management & Science, Jhinga, WB during the academic year 2016-2017.

#### *Signature of Supervisor(s)*

#### 

#### 

#### Name & Designation

Project Supervisor(s)

***Head of the Department***

Department of Computer Science & Engineering

Neotia Institute of Technology, Management & Science, Jhinga, West Bengal

Date:

Place:

**ACKNOWLEDGEMENT**



Acknowledgement should be prepared by their wordings showing their gratitude towards department.

**Anisha Pal**

**ABSTRACT**

Data clustering is an unsupervised data analysis and data mining technique, which offers reﬁned and more abstract views to the inherent structure of a data set by partitioning it into a number of disjoint or overlapping (fuzzy) clusters. Hundreds of clustering algorithms have been developed by researchers from a number of different scientiﬁc disciplines. The intention of this report is to present a special class of clustering algorithms, namely K-means and K-medoids (under partitioning based). After the introduction and a review on iterative relocation clustering algorithms, a proper implementation of cluster analysis is presented. Cluster validation is an important step in cluster analysis; it is the evaluation of similarity between two clusters.  Cluster analysis are used everywhere be it during public census, classification of species of animals and plants or similar diagnostic cases     . To run a good business, classification and clustering your work is as important as taking other managerial decisions