National Institute of Technology Rourkela

Departmental Seminar

Seminar Title	: A Multi-Step Fuzzy C-Means Approach for Accurate Data Imputation in Healthcare
Speaker	: Subhashish Nayak (520cs2008)
Supervisor	: Prof. S. Pyne, PIC, Seminars
Venue	: Convention room CSE Dept.
Date and Time	: 30 Oct 2024 (11:00 AM)

Abstract : In this emerging technological era, data is the new oil. For a long time, missing values in data posed a huge challenge to machine learning, data statistics, data mining and other datadriven fields. In the present context, various data imputation methods to handle missing data exist as discovering meaningful information is essential. However, the most widely used approach to handle missing values in a huge dataset is to discard those values, leading to losing crucial information. Therefore, a novel imputation method needs to handle those missing values. Soft clustering-based approaches are widely employed in many current data imputation techniques applications. This paper proposes an accurate Fuzzy C-Means (FCM) clustering and integrates it with membership values for weighted imputation. The contributions include a novel methodology for estimating missing values in healthcare datasets, retaining the dataset's underlying distribution while maintaining vital information, proposed workflow, and handling numerical and categorical data types. This multi-step procedure yielded more accurate results and representative information than other state-of-the-art methods: Mean imputation and Fuzzy C-means with Genetic Algorithm (FCMGA). The proposed algorithm outperforms the available methods and is presented in this work. The experimentation is carried out on two benchmark datasets to assess the efficacy of the proposed approach. The proposed method gave significantly improved MSE, NRMSE, UCE and CCD scores on Diabetes and Heart datasets.