
Departmental Seminar

Seminar Title	: Reinforcing Clustering Accuracy in Healthcare Data Through Cluster-Driven Matrix Factorization for Accurate Data Imputation
Speaker	: Subhashish Nayak (520cs2008)
Supervisor	: Sumanta Pyne, PIC Seminar, CSE
Venue	: Convention room CSE Dept.
Date and Time	: 03 Mar 2025 (11:00 AM)
Abstract	: People are more concerned about their health these days, and the accuracy of medical information has come into focus. Since medical data frequently contains missing values, the field of medical data imputation has recently seen a surge in activity. This work presents a method that combines data imputation and clustering optimization to handle these challenges. Missing values are filled using matrix factorization with Truncated Singular Value Decomposition (SVD), which estimates the missing entries while preserving the overall structure of the dataset. The sensitivity of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to parameter values, which can result in ineffective cluster detection, is one of the shortcomings of the current imputation techniques highlighted in this work. For clustering, the method automatically optimizes DBSCAN parameters (eps and min samples) using the silhouette score to improve cluster quality. The suggested approach is evaluated using generic metrics like Mean Squared Error (MSE) and Mean Absolute Error (MAE) on benchmark datasets with varying degrees of missing data (5% to 30%). The method ensures accurate clustering results while lowering imputation errors, according to the results. This approach offers a straight- forward but efficient way to deal with missing data in medical application.