
Seminar Title	: Detection of Anomalies on non-structured High Dimensional Data using Unsupervised Approaches
Speaker	: Satyanarayanamurty M (Rollno : 920cs5009)
Supervisor	: Prof. Ramesh Kumar Mohapatra
Venue	: CS 323, 2nd Floor, CS Department
Date and Time	: 02 May 2024 (17:00 hrs)
Abstract	: Outlier detection plays a crucial role in identifying rare and potentially anomalous data points in high-dimensional datasets. The challenges posed by high-dimensional data necessitate the development of specialized techniques for dimensionality reduction, feature selection, and data analysis to extract meaningful patterns and insights. Recently, deep learning techniques, specifically autoencoders, have gained attention for their ability to minimize reconstruction errors and identify outliers based on higher reconstruction errors. Recent work leverage, Density Peak Clustering (DPC) and Self Organizing Map (SOM) as clustering approaches to identify probable outlier points based on density and distance to higher density points. While DPC has the drawback of setting a threshold for density, SOM addresses this limitation. But SOM doesn't fit well for large scale high dimensional data and more training is needed. However, autoencoder-based models often overestimate reconstruction errors for normal points and underestimate them for outliers, leading to the potential overlooking of genuine outliers. proposes a novel approach that combines the Reverse Density Peak Clustering Approach (RDPOD) with Extreme Learning Machines (ELM) to enhance outlier detection accuracy. RDPOD is employed as a clustering algorithm to identify density peaks and partition the data into clusters based on local density information. Subsequently, the outlier scores generated by RDPOD are used as input features for an ELM classifier. ELM leverages the nonlinear mapping capabilities to learn the complex relationships between the input features and the outlier class, enabling robust outlier detection in high-dimensional spaces. Experimental evaluations on benchmark datasets demonstrate the effectiveness of the proposed RDPOD-ELM framework in accurately identifying outliers while maintaining computational efficiency.