**Abstract**

Data mining is the process of finding the previously unknown and potentially interesting patterns and relation in database. Data mining is the step in the knowledge discovery in database process (KDD). The structures that are the outcome of the data mining process must meet certain condition so that these can be considered as knowledge. These conditions are validity, understandability, utility, novelty, interestingness. Researcher identifies two fundamental goals of data mining: prediction and description. The proposed research work suggests the semi-supervised clustering problem where to know (with varying degree of certainty) that some sample pairs are (or are not) in the same class. A probabilistic model for semi-supervised clustering based on Shared Semi-supervised Neighbor clustering (SSNC) that provides a principled framework for incorporating supervision into prototype-based clustering. Semi-supervised clustering that combines the constraint-based and fitness-based approaches in a unified model. The proposed method first divides the Constraint- sensitive assignment of instances to clusters, where points are assigned to clusters so that the overall distortion of the points from the cluster centroids is minimized, while a minimum number of must-link and cannot-link constraints are violated. Experimental results across UCL Machine learning semi-supervised dataset results show that the proposed method has higher F-Measures than many existing Semi-Supervised Clustering methods.