

### Assignment 3

1. Implement the following Feature selection Algorithm (PCFS):
  - i) Find the Pearson correlation coefficient (pc) between every pair of features.
  - ii) Construct a graph with features as nodes. Assign pc between two features as the weight of the edge between two nodes corresponding to two features.
  - iii) Remove all the edges with weight less than a predetermined threshold.
  - iv) Compute the importance of a node as the sum of weights of all edges incident on it.
  - v) Select the most important node and remove it from the graph.
  - vi) Repeat step (iv) and (v) until the graph becomes a null graph
  - vii) The nodes selected are considered as the subset of features.
2. Install Weka tool:

<http://www.cs.waikato.ac.nz/ml/weka/citing.html> in your computer and learn it using help provided in it.

  - i) Consider at least five datasets available in UCI Machine Learning Repository and apply inbuilt various preprocessing, techniques. Make a table to compare the results.
  - ii) Let your feature selection algorithm is PCFS. Use at least six (such as InfoGain, Classifier subsetEval, Correlation AttributeEval, ReliefF AttributeEval etc.) feature selection algorithms available in weka tool.
  - iii) Create all the reduced datasets and find the classification performance (accuracy, precision, recall, f\_score) for all datasets using ten classifiers (like ID3, C4.5, SVM, Naïve Bayes, Neural Network, KNN, Bagging, Boosting etc), one from each type. Use 10-fold cross validation technique while applying the classifiers.

**Note:** Some feature selection algorithms in weka provide reduced feature subset from original feature set. Some others rank all the features of the feature set. In these cases, considered first P number of features, where P is the number of features selected by your algorithm.