**Assignment 3**

**Clustering**

*Data Mining Fall 2021*

**Due Date:13 December 2021**

In this assignment, you have to pre-process and cluster the dataset provided to you in Assignment1 **using Python. You have to submit your python code along with a word document explaining your results and findings.**

1. Pre-Process the dataset using different techniques like normalization, discretization, concept hierarchy, correlation using Python. Explain which pre-processing steps are performed on the data and **why?**
2. Perform **K-MEANS** clustering on the given dataset. Try to formulate a question for which you wish to cluster the dataset. Select a subset of attributes that can be beneficial for your task *. Don’t use all the attributes.* 
   1. Run K-means multiple times for each K. Report your findings (error in each clustering, the time required)
   2. Run K-means with **different K**
   3. Report the K that gives the best result for each subset of attributes.
   4. Indicate the number of iterations to convergence for different runs.
   5. Examine the quality of clusters and also of clusterings. Report the errors: within-cluster sum of squared error and between cluster sum of the square error for each run of K-mean.
3. Cluster the dataset using **Hierarchical Clustering (single link, complete link, average link).** 
   1. Run for different values of the number of clusters. Include the dendrogram and time taken by each clustering in your report.
   2. Examine the quality of clusterings. Report the errors.
4. Draw different plots to visualize the clustering results. (include plots in your report).
5. ***Compare the clustering results of the K-means and Hierarchical in terms of time and quality of clustering.***