# Data Mining Lab Assignment

CS524

August 27, 2021

## Instructions:

- You can use any programming language of your choice with the suitable input format for each of the question (Python will be preferable). Each question should have a readme file and the detailed report.
- A readme file should precisely tell how to compile and run your program. Give the exact commands with respect to the datasets provided.
- The marks will be given on the basis of quality of code, use of innovative data structures, scalability, correctness, and completeness of the report. Quality of code includes proper comments wherever necessary and proper code indentation. Please submit your code along with the readme file and the report.
- You are supposed to submit the assignment on google classroom no later than **17th November 2020**. This is a strict deadline and any assignment submitted later will not be consider for evaluation unless you take prior permission (at least 4 days before the submission deadline).
- Submit the zip file with naming convention as DM_entrynumber_assignno.zip eg. DM_2020csz0004_assign1.zip
- Follow code of conduct strictly. If code is found to be copied then we are entitled to give you zero for the entire assignment.

## Problem 1:

Download the Libras Movement data set. Apply PCA to the data set, and report the eigenvectors and eigenvalues.
(https://archive.ics.uci.edu/ml/datasets/Libras+Movement) [15 Marks]

## Problem 1:

Implement the three frequent itemset mining algorithms: [30 Marks]

1. Apriori Algorithm
2. FP-growth
3. Eclat algorithm

Compare the performance of each algorithm on the below datasets:

1. T20i6D100k Dataset (https://www.philippe-fournier-viger.com/spmf/datasets/t20i6d100k.txt)
2. BMS WebView 2 (KDD CUP 2000) (https://www.philippe-fournier-viger.com/spmf/datasets/BMS2_itemset_mining.txt)
3. Chess (https://www.philippe-fournier-viger.com/spmf/datasets/chess.txt)
4. Liquor 11 (https://www.philippe-fournier-viger.com/spmf/liquor_11frequent.txt)

Prepare a detailed report containing the following:

- Details of the algorithms implemented. Did you use any optimization techniques on the top of the original algorithm. If yes, explain them.
- Give a brief description of each of the datasets such as how many transactions it had, average width of the transaction, size of maximal frequent itemsets, number of maximal frequent itemsets, total number of items, etc.
- Comparison of the three algorithms in terms of time and space complexity on different datasets.
- Explain why you think a particular algorithm works better in one dataset as compared to other.

.