**Assignment 3 Data Mining CSE 572**

Given: Meal Data of 2 subjects

Amount of carbohydrates in each meal

Todo:

1. Extract features from Meal data
2. Cluster Meal data based on the amount of carbohydrates in each meal

Extracting Ground Truth:

Derive the max and min value of meal intake amount from the Y column of the Insulin data. Discretize the meal amount in bins of size 20. Consider each row in the meal data matrix that you generated in Assignment 2. According to their meal amount label put them in the respective bins.

In total you should have bins.

Performing clustering:

Now ignore the mealAmountData. Without using the meal amount data use the features in your assignment 2 to cluster the mealDataX.csv into clusters. Use DBSCAN or KMeans. Try these two.

Report your accuracy of clustering based on SSE, entropy and purity metrics.

Grading:

50 points for developing a code in Python or Matlab that takes the dataset and performs clustering

20 points for developing a code in Python or Matlab that implements a function to compute SSE and SSE, entropy and purity metrics. These two can be written in the same file.

30 points will be evaluated on the supervised cluster validation results obtained by your code. This will be compared against class average to determine the final score.