Consider the data from Project 1 (feature-extraction) .

Step 1: Extract the bolus insulin data from InsulinBolusLunchPartX.csv. This is the maximum insulin level in the vector. I_B

Step 2: Find the maximum CGM from the CGMSeriesLunchPartX.csv file. CGM_M (Quantize, divide CGM

values into bins, of size 10mg/dL. ~50 – 60, 60 – 70, 70 – 80, …… 350)

Step 3: Find the CGM value that is at the time when the lunch was taken. I.e. the sixth sample. CGM_0 (Quantize, divide CGM values into bins, of size 10mg/dL. ~50 – 60, 60 – 70, 70 – 80, …… 350)

Extract these metrics from all lunch instances for all subjects.

1. Report the most frequent itemsets for each of the subjects (Bin for CGM_M, Bin for CGM_0, Insulin Bolus)
2. Consider the rule of the form: $\{Bin\ for\ CGM_M, Bin\ for\ CGM_0\} \rightarrow I_B$

   Find the rule with the largest confidence for each subject.

   Extract all rules that you observe. Calculate confidence of each observed rule.

3. Find anomalous events by finding the least confidence rules.
   Rank rules according to confidence.

   Expected Output
   1. CSV File with most frequent sets. One row for each itemset.
   2. CSV file with largest confidence rules. One row for each rule. Rules are of the form $\{Bin\ for\ CGM_M, Bin\ for\ CGM_0\} \rightarrow I_B$
   3. Anomalous rules, Rules with confidence less than 15 %. One row for each rule.

      Grading:

      If your code returns three CSV files with rules. Then we code will be analyzed to see if you did the rule search. (70 points)

      Majority rules. If your rule falls in the majority set your get 30
      If x percent of your rules fall in majority set you get $p = 30 * \frac{x}{100}$