CIS635 team project – CDC Diabetes

List of tasks

Presentation of the data – May 22, 8PM

1. Get the dataset and isolate the target variable and the features used for the study (6 columns)
2. Look for formatting needs, data binning, discretization
3. Descriptive statistics, data charts

Algorithm 1: Association rules using Python - May 25, 10AM

1. Running the codes on Colab
2. Link with Github?
3. Present findings
4. Interpreting the results

Algorithm 2: Logistic Regression using R

1. Running the codes on R Studio
2. Link with Github
3. Present findings
4. Interpreting the results

If there is no significant findings or interpretation,

A larger set of features or all the features

1. Use the most appropriate method
2. Running the codes on Colab
3. Link with Github?
4. Present findings
5. Interpreting the results