**Report**

**Abraca-data**

**Date:** 2nd March 2022

**Tasks performed in this week**

1. Multivariate analysis
2. Feature Selection

**Outcomes of the tasks performed**

1. 3 different variables are compared with age. cigsPerDay, totChol and glucose are compared with age. These parameters are most correlated with the 10 year prediction of CHD.
2. Then graphs are plotted to understand the relationship between systolic blood pressure and diastolic blood pressure for patients based on their gender and whether they are current smokers or not and plots the best fit line. A basic regression plot is plotted using the sns.lmplot function of the seaborn library.
3. Identified the top 10 features which contribute highly to ten year CHD using Chi-Score value. The features such as education, BMI and current smoker seemed to be having low chi-score value and hence these are to be ignored in later analysis.

**Tasks to be performed in the upcoming week**

Further, we have planned to perform dimensionality reduction on our data.

We will then apply logistic regression and KNN on the data on which we have performed dimensionality reduction.