My Project

**Name : Ashtageeeta Ashok Hingmire**

**CWID : 50333659**

Q1: Build a logistic regression model, and use all given independent variables to predict the probability that a patient has type 2 diabetes. In your assignment submission, please make a screen capture of your analysis results. Please indicate which independent variables are significant, and the practical implications if they are significant.

**Rcode snap**

Text

Description automatically generated

***Formula :***

**Pr(Yes) = 1/1+e-y**

So according to above R code

Y= -8.33+0.005\*BodyWeight+0.278\*GymVisits+0.630\*StateMD+(-0.905)\*StateND+(-0.485)\*StateNY+(-0.812)\*StateSD+(-0.253)\*StateTX+(-.849)\*StateWA+0.212\*SugarIntake+0.027\*ChLevel

**Example**:

A person with a BodyWeight 102.5 visits gym 2 times and is from MD and has a SugarIntake of 4.2

and ChLevel of 150

Y= -8.33+0.005\*102.5+0.278\*2+0.630+0.212\*4.2+0.027\*150

Y= -1.6911

Probability of this person having 2 diabetic

Pr=1/(1+e-(-1.6911))=0.155

From the above R sinppet, the independent variables which are significant are – ChLevel and GymVisits as the p-value for both these variables are less than or equal to 0.05.The remaining independent variables have p-value greater than 0.05 and hence, are considered as non-significant variables.

The practical implication we can deduct is that if we take any person in the population and get to know their ChLevel and their GymVisits we will be able to predict the probability of them having the Type2 Diabetes.

Q2: Build a decision tree model, and use all given independent variables to predict the probability that a patient has type 2 diabetes. In your project submission, please make a screen capture of your analysis results. Also, create a decision tree diagram to visualize your analysis.

**Rcode snap**

Text

Description automatically generated

**Decision Tree**

Timeline

Description automatically generated