

# Machine Learning Report

*Keith Engwall*

*4/20/2018*

## **Applying Machine Learning to Project**

How well can patient data predict the occurrence of diabetes in a patient?

This is a supervised problem because we are attempting to predict a specific dependent variable using a model based on a set of independent variables.

This is a combination of classification and regression.

Based on some model testing, the independent variables I will use are Gender, Age, AllergyType and Diagnosis Category.

Because the independent variables are a mix of discrete/categorical variables and continuous variables, the Classification and Regression Trees (CART), specifically Random Forests, type model is a good choice.

I will be testing the random forest model both using a train/test split in the original data, and also using a new test data set.