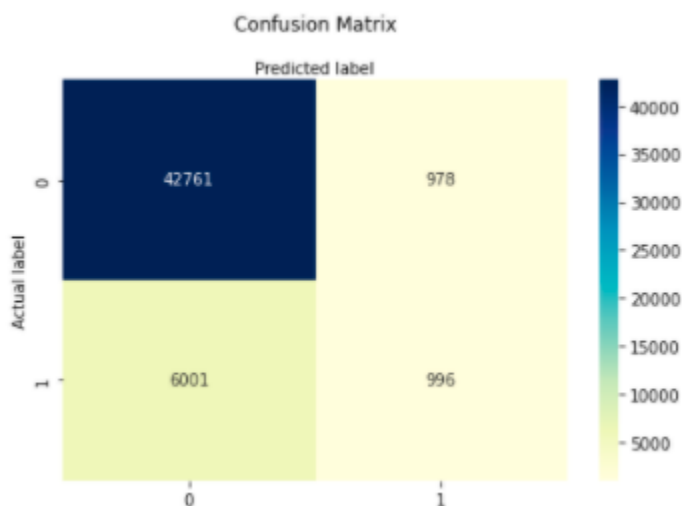


Goal: I plan to build a system that analyzes the vitals of individuals and will determine whether they have diabetes or not. A model with acceptable accuracy, recall, precision and F1 will be utilized for a more accurate presentation.

Process: The data set leveraged in this analysis was obtained via [kaggle](https://www.kaggle.com). It contains over 250,000 observations and 10 features.

Preliminary conclusions: I created a logistic regression, decision tree and random forest algorithm to analyze my data. Right now the logistic regression has the highest accuracy. I will perform cross validation and fine tune my models and will determine which is the best approach to use for my model.

Logistic Regression



Accuracy: 0.8624448123620309
Precision: 0.8769328575530126
Recall: 0.9776400923660806

Feature: 0, Score: 0.16320
Feature: 1, Score: -0.64301
Feature: 2, Score: 0.48497
Feature: 3, Score: -0.46190
Feature: 4, Score: 0.18432
Feature: 5, Score: -0.11978
Feature: 6, Score: -0.40602
Feature: 7, Score: 0.03772
Feature: 8, Score: -0.51785
Feature: 9, Score: 0.26540

