

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one in front of the green one.

Predicting Insulin Prescriptions

Using the CDC's NHANES Study



The Dataset

Health information of approximately 5,000 people, including 168 questions regarding diet and a list of medications that each participant takes.

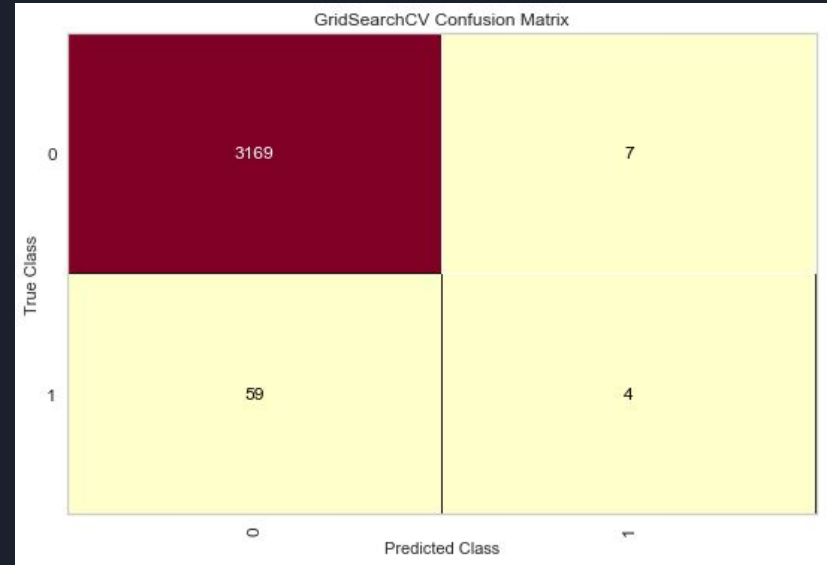
The data was cleaned and formatted into a new dataframe with a True/False format for our target variable, Insulin.

Predicting Insulin with All Features

The top features for predicting insulin prescriptions:

Special diets, Sugar Consumption, Caffeine Consumption

Would recommend evaluating the diet of those patients that the model falsely predicted an insulin prescription

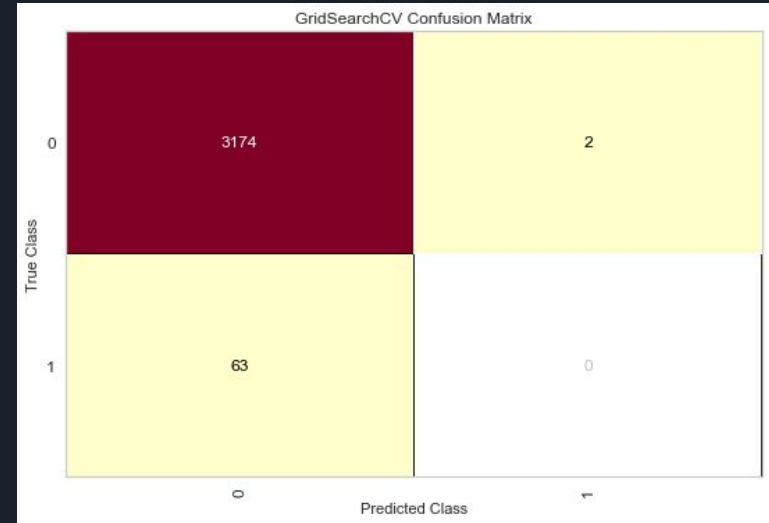


The Effect of Removing 'Cheat' Features

14 features relating to special diets and sugar consumption were removed for further exploration.

Even after optimizing model, the F1 score is 0.0

We cannot reliably look at any feature as an indicator of insulin prescriptions.

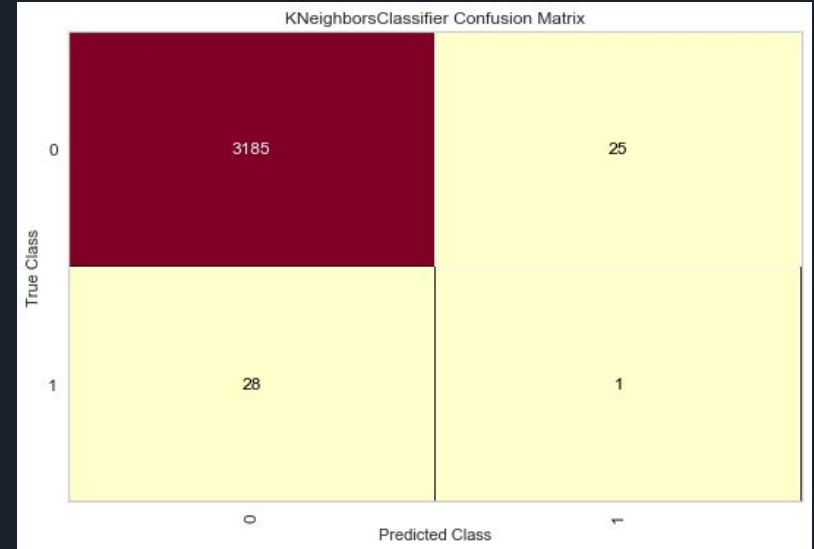


Predicting Glipizide Prescriptions

Because insulin is prescribed to both Type 1 and Type 2 diabetics, Glipizide (a popular anti-diabetic drug) was also investigated

The dataset was even more unbalanced than insulin

There are 25 patients who may benefit from dietary changes





Ideas for Future Work

Balance data using SMOTE

Determine feature importances of KNN Glipizide model

Investigate other ways to solve the same problem- blood insulin levels, diagnosis codes



Conclusion

Any questions?