Hudson Pak, Justin Lewinski, Everett Prussak
MGSC 310 - Section 2 Team 9

Students must upload to Canvas a **one-page** outline of their project. This outline should include

a) identify a dataset you will use

We will be using:
https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset

b) the outcome you are trying to predict, and what variables you will use to predict it;

We are predicting whether a patient has diabetes based on certain diagnostic measurements. We will use the following variables:

Pregnancies: To express the Number of pregnancies

Glucose: To express the Glucose level in blood

BloodPressure: To express the Blood pressure measurement

SkinThickness: To express the thickness of the skin

Insulin: To express the Insulin level in blood

BMI: To express the Body mass index

?? DiabetesPedigreeFunction: To express the Diabetes percentage

Age: To express the age

Outcome: To express the final result 1 is Yes and 0 is No

c) motivation to your project -- as in the business or practical management use case of such a prediction;

Diabetes is as prevalent as ever in America and is one of the biggest problems impacting the youth and adult population of the country. We are extremely interested in

running the numbers and getting to the bottom of what causes it, which can lead to preventative measures in the future.

d) two methods you will use to analyze your question of interest;

The first method we will use to analyze is using a 80/20 Train-Test-Split with Logistic Regression.

The next method we will use to analyze is using K-Fold Logistic Regression. We will have to discuss further on how we choose the hyperparameter K. Since our data is 900 rows we felt that it was too large to use LOOCV.

e) the names of the students who will be part of your group.

Hudson Pak, Justin Lewinsky, Everett Prussak

f) Rmarkdown document showing the summary statistics against the baseline dataset.

We added the RMD and HTML of the summary statistics and a head of the file.

If you have difficulties identifying a dataset or project please reach out to the professor. If you have a dataset from an internship, consulting opportunity or job you have been meaning to analyze, you are welcome to use this dataset. After reviewing your projects I may suggest alterations to the project.