**Diabetics Prediction Using Logistic Regression**

1. **Dataset**
   1. **Data Source**

The data being used for this study was made available by “National Institute of Diabetes and Digestive and Kidney Diseases” as part of the Pima Indians Diabetes Database. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here belong to the Pima Indian heritage (subgroup of Native Americans), and are females of ages 21 and above. This dataset was available in Kaggle dataset as an open dataset for study. The dataset has a total of 8 features to determine the final outcome i.e. whether the concerned person is diabetic or not. For this purpose a total of 768 samples being considered. Link for the complete dataset in GitHub repository is as follows:

https://github.com/sourinsett/Capstone\_Coursera/blob/main/diabetes2.csv