

REPORT - Lab 14 Naive Bayes' Classifier

Vrushali Pandit B20BB047

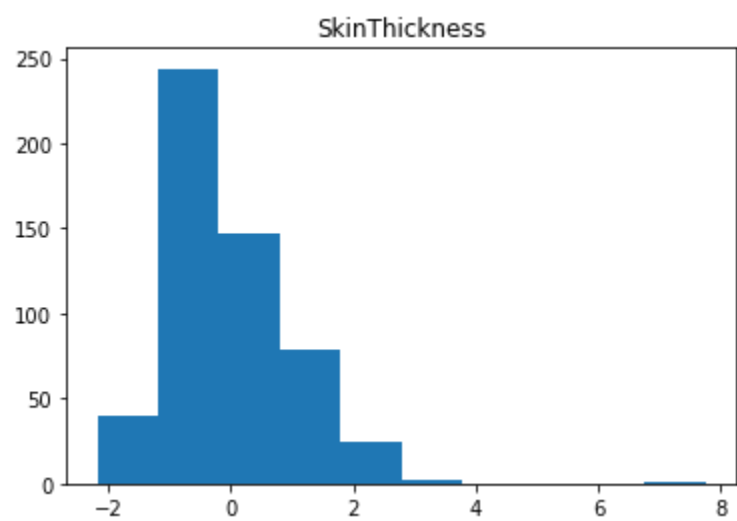
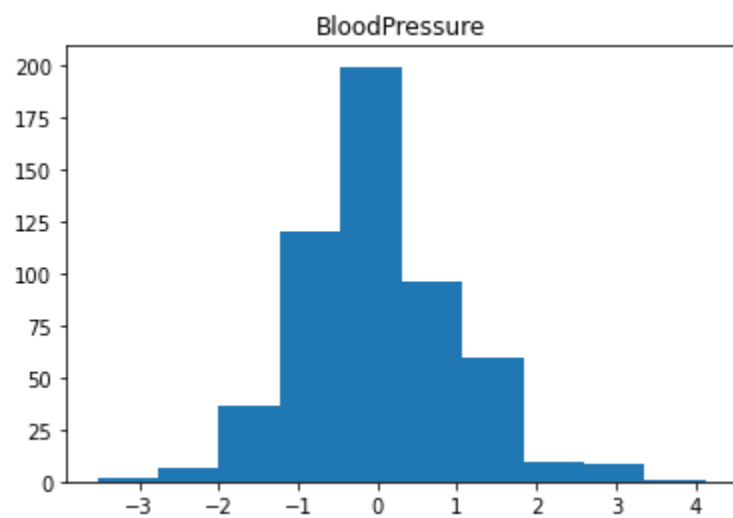
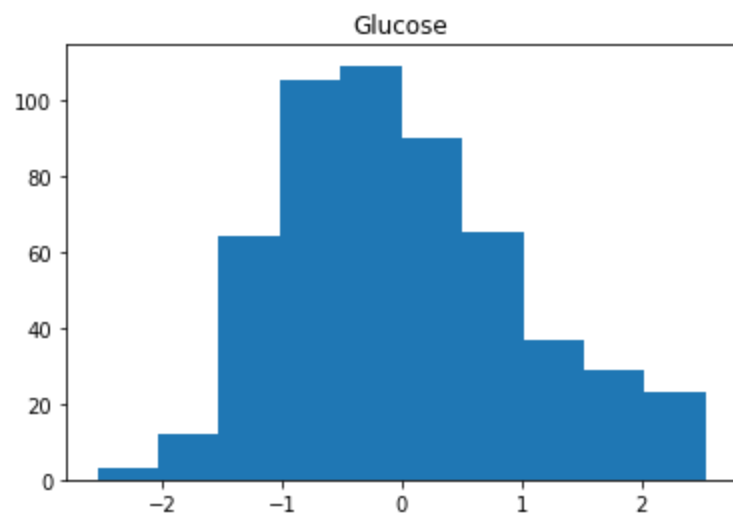
https://drive.google.com/file/d/1YACFmhlHseOHeSU7_umo_47NrMA8q77L/view?usp=sharing

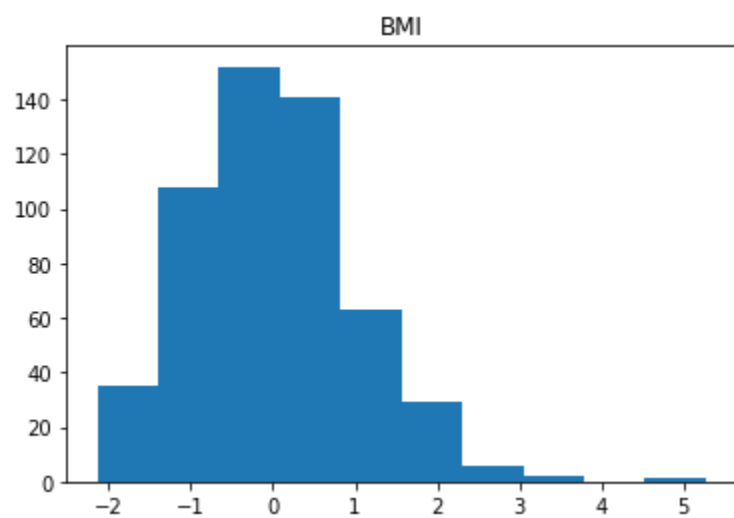
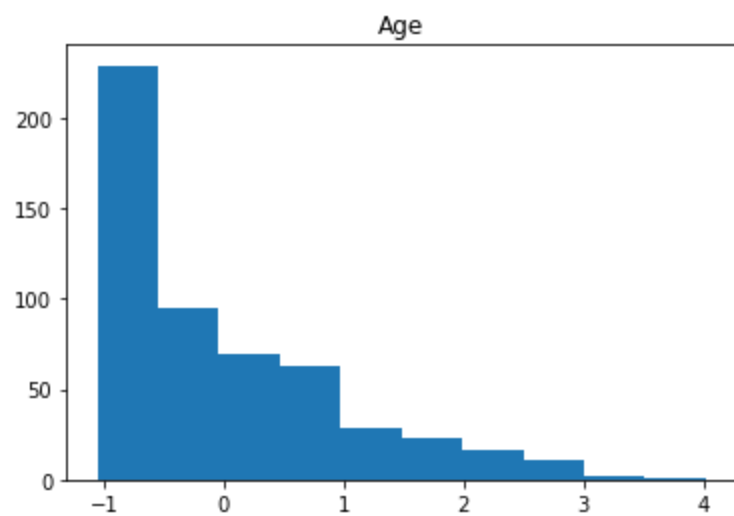
Data:

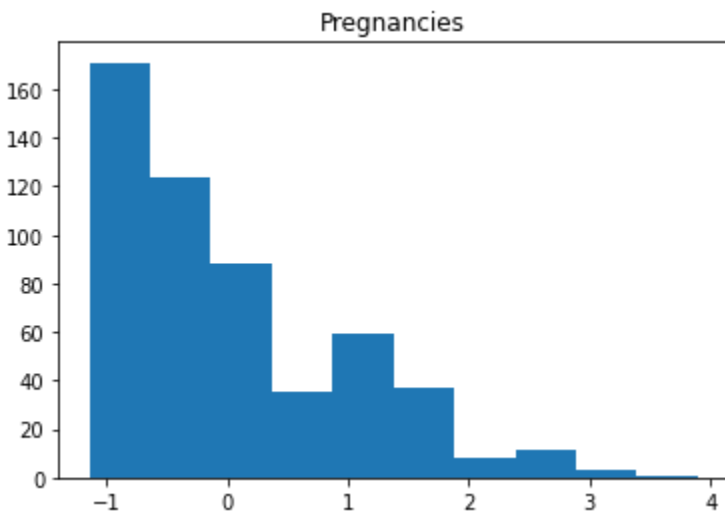
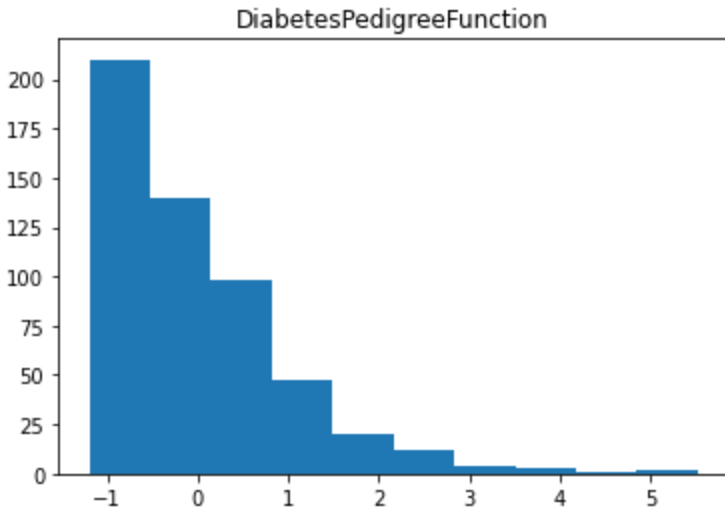
The dataset is data about people with or without diabetes accounting the features which are thought to be correlated with diabetes such as the number of pregnancies, glucose level, age etc.

Pre-processing:

- Many features (Glucose, BloodPressure, SkinThickness, BMI) have 0 values which don't make sense. These are imputed by their means.
- Used standard scaling to ensure that one of two assumptions of naive bayes' classifier is held true (i.e. all features have equal contribution).
- To observe what kind of data we are dealing with, after scaling, I performed histogram plotting for all the features.
 - We observe that some are gaussian features, some are skewed, while some are random distributions.







Model Building

Using the sklearn library we get an accuracy of 0.732, while using Naive Bayes' classifier built from scratch defined for individual features we get accuracies of 0.64 by using pregnancy and an accuracy of 0.718 using Glucose levels.