

# Diabetes

By:  
Amwaj Hantool  
Meshari Abdullah  
Abdulwahab Alsubit

# Out Line



**Introduction**

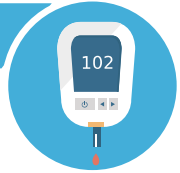


**Work Flow**



**visualization**





This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

The Dataset contain

10 columns

767 rows

Pregnancies, Glucose, BloodPressure,  
SkinThickness, Insulin, Gander, Age, Outcome,  
DiabetesPedigreeFunction and BMI

# Work Flow

Import Lib

Data Cleaning

Data visualization

Data Pre-Processing

Baseline Model

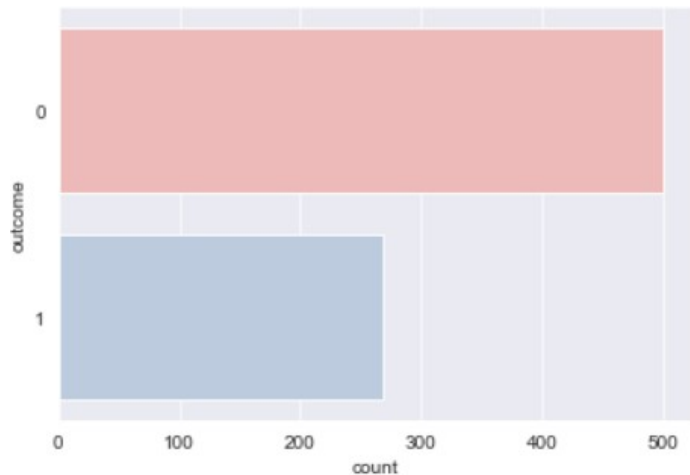
Bagging Algorithms

Boosting Algorithms

Random Forest: 100.00 %

XGBoost: 100.00 %

# Data visualization



Base on chart data is not balance

