Idea 1: Classify breast cancer using breast measurements taken from patients

The dataset was collected by University of Wisconsin Hospitals. Measurements include: radius, texture, perimeter, area, and smoothness. Each patient either received a diagnosis of breast cancer (marked as 1) or no diagnosis (marked as 0). The goal is to classify undiagnosed patients with a model.

<https://www.kaggle.com/merishnasuwal/breast-cancer-prediction-dataset>

Idea 2: Predict diabetes using a more accurate model than logistic prediction

The dataset was collected by the National Institute of Diabetes and Kidney Diseases and restricted to those with Pima Indian heritage. They only included females that are 21 and above in this study. Some of the collected information included: number of pregnancies, glucose level, blood pressure, insulin, age, and BMI. The goal is to predict if a patient will get diabetes based on current health information.

<https://www.kaggle.com/kandij/diabetes-dataset>

Idea 3: Classify patients as inpatient or outpatient based on blood test lab results

The dataset was collected by a private hospital in Indonesia. The dataset reads like a typical blood work result from the lab with patient names hidden. The goal is to classify patients waiting to be admitted or released.

<https://www.kaggle.com/manishkc06/patient-treatment-classification>