

This document gives brief overview of the business problem and the programming approach to solve the unsupervised learning problem here.

The aim here is to classify the cell nuclei as Benign (non-cancerous) or Malignant (cancerous) based on the data features collected by Wisconsin Breast Cancer dataset. It contains the average values of about 20 data points (columns) describing the instances of actual diagnosis. We treat this problem as unsupervised learning taking out the diagnosis information.

As a part of clustering approach, we employed a couple of techniques here namely Hierarchical Clustering and KMeans clustering. We also explored the Principal Component Analysis technique and visualized how the feature reduction technique can be exploited. We compared the table plots for comparing how the instances are grouped under clusters across various approaches.