Assignment 1

There are 3 types of datasets given below. The details of these three datasets are:

a) Linearly separable: These are 2-dimensional artificial data of 2 classes that are linearly separable. Each class has 1000 data points.

b) Non-linearly separable : These are 2-dimensional artificial data of 2 classes that are non-linearly separable. Each class has 1000 data points.

c) Real-world: Real-world data of 3 classes. (Number of data points for class1 = 2388, class2 = 2164 & for class3 = 2291)

Perform

1. Classification using Bayesian classifier (including parameter estimation and decision)

2. Classification (only for the 2 artificial datasets) when you approximate the pdfs via normalized histograms (considering independence between features - thus, involving 1D histogram computations for each of the two features)

In both cases, do the following:

For each task, you can divide the data into training and testing parts.
Build the model with the training data and evaluate its performance using the test data.
A typical training testing ratio can be 80% - 20%
Make a confusion matrix to show your results.

For the artificial datasets, also show a 2D scatter plot for the training data and test data with different colours. Also, points for different classes should also be shown in different colours.