Datasets:
https://archive.ics.uci.edu/ml/datasets.html?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=table

https://machinelearningmastery.com/standard-machine-learning-datasets/

https://archive.ics.uci.edu/ml/datasets.html?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list

1. **Iris dataset**
    a. Link: https://archive.ics.uci.edu/ml/datasets/Iris
    b. Des: This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

       Predicted attribute: class of iris plant.

       This is an exceedingly simple domain.

       This data differs from the data presented in Fishers article (identified by Steve Chadwick, spchadwick '@' espeedaz.net). The 35th sample should be: 4.9,3.1,1.5,0.2,"Iris-setosa" where the error is in the fourth feature. The 38th sample: 4.9,3.6,1.4,0.1,"Iris-setosa" where the errors are in the second and third features.

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica

2. **Lung cancer**
    a. Link: https://archive.ics.uci.edu/ml/datasets/Lung+Cancer
    b. Des: This data was used by Hong and Young to illustrate the power of the optimal discriminant plane even in ill-posed settings. Applying the KNN method in the resulting plane gave 77% accuracy. However, these results are strongly biased (See Aeberhard's second ref. above, or email to stefan '@' coral.cs.jcu.edu.au). Results obtained by Aeberhard et al. are :

RDA : 62.5%, KNN 53.1%, Opt. Disc. Plane 59.4%

The data described 3 types of pathological lung cancers. The Authors give no information on the individual variables nor on where the data was originally used.

Notes:
- In the original data 4 values for the fifth attribute were -1. These values have been changed to ? (unknown). (*)
- In the original data 1 value for the 39 attribute was 4. This value has been changed to ? (unknown). (*)

Attribute Information:

Attribute 1 is the class label.

All predictive attributes are nominal, taking on integer values 0-3

3. **Face Images:**
   a. https://archive.ics.uci.edu/ml/datasets/CMU+Face+Images
   b. Des: Each image can be characterized by the pose, expression, eyes, and size. There are 32 images for each person capturing every combination of features.

      To view the images, you can use the program xv.

      The image data can be found in /faces. This directory contains 20 subdirectories, one for each person, named by userid. Each of these directories contains several different face images of the same person.

      You will be interested in the images with the following naming convention:

      .pgm
      is the user id of the person in the image, and this field has 20 values: an2i, at33, boland, bpm, ch4f, cheyer, choon, danieln, glickman, karyadi, kawamura, kk49, megak, mitchell, night, phoebe, saavik, steffi, sz24, and tammo.
      is the head position of the person, and this field has 4 values: straight, left, right, up.
      is the facial expression of the person, and this field has 4 values: neutral, happy, sad, angry.
      is the eye state of the person, and this field has 2 values: open, sunglasses.
      is the scale of the image, and this field has 3 values: 1, 2, and 4. 1 indicates a full-resolution image (128 columns by 120 rows); 2 indicates a half-resolution image (64 by 60); 4 indicates a quarter-resolution image (32 by 30).
      If you've been looking closely in the image directories, you may notice that some images have a .bad suffix rather than the .pgm suffix. As it turns out, 16 of the 640 images taken have glitches due to problems with the camera setup; these are the .bad images. Some people had more glitches than others, but everyone

who got ``faced'' should have at least 28 good face images (out of the 32 variations possible, discounting scale).

4. Breast Cancer:

a. Link: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer

b. Des:

This is one of three domains provided by the Oncology Institute that has repeatedly appeared in the machine learning literature. (See also lymphography and primary-tumor.)

This data set includes 201 instances of one class and 85 instances of another class. The instances are described by 9 attributes, some of which are linear and some are nominal.

Attribute Information:

1. Class: no-recurrence-events, recurrence-events
2. age: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99.
3. menopause: lt40, ge40, premeno.
4. tumor-size: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59.
5. inv-nodes: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39.
6. node-caps: yes, no.
7. deg-malig: 1, 2, 3.
8. breast: left, right.
9. breast-quad: left-up, left-low, right-up, right-low, central.
10. irradiat: yes, no.

Requirements:

1. Run k-NN, Decision Tree, Random Forest and SVM for each dataset
2. Show the Precision, Recall and Accuracy for each dataset.
3. With SVM, change the Kernal and see the differences.