Bank Note Authentication

In this project, you will require to use the data set named as: ***data_banknote_authentication.csv***. You will load the data using ***pandas*** library. Additionally, you will require to use the following packages.

- ***numpy***
- ***matplotlib***
- ***knn_numpy***

Details on dataset can be found:
https://archive.ics.uci.edu/ml/datasets/banknote+authentication
https://drive.google.com/file/d/18m_lBOVp37BOH6osFmux6VPS1S_WfXoX/view?usp=sharing

Step 1: Pre-processing:

In this step you must get features and labels (classes) of the dataset.

Step 2: Training the model:

Using ***train_test_split*** method from ***sklearn*** you will split data into 2 parts. One used for training a classifier. Another to test how generalized is our classifier.

Case-1: use 20% (0.2) of data for testing; Case-2: use 35% (0.35) of data for testing

Step 3 Scale data:

As you can see data in the dataset is not on the same scale (some data is -13 some data between 0 and 1), for better performance/prediction you need to scale all data into the same scale.

Step 4. Prediction:

- Show the Predicition using custom KNN
- Show the Prediction using Sklearn KNN

Step 5: Test more deeply sklearn algo:

You need to use 2 more params to test accuracy for Sklearn algorithm. Precision and Recalls

- Precision - Scoring measurement which is calculating how many positive predictions that are correct
- Recalls - Scoring measurement used to evaluate how to classifier is

Note: Precision and Recalls are note the best choice to evaluate classifier. There are other metrics as well such as F1 and ROC AUC

Step 6: generate necessary graphs for visualizations using seaborn and matplotlib packages.

Your report must include the followings:

a) Background of the problem
b) Detail description of Datasets (Including attributes, data types, values etc.). You must reflect your understanding of the dataset.
c) Detail explanation of KNN algorithm
d) Implementation of the case in python. (Must include code in the appendix section)
e) Analysis of your result: Must explain any graphs, plots.
f) Discussion: Overall discussion on the problem and the implemented solution.