K-Nearest Neighbor Implementation

You need to complete the following tasks in this assignment:

1. Load Data into numpy array: **10%**
    a. Iris Dataset: iris.csv
    b. Digits Dataset: train.csv
2. Split Dataset to Train/Validation/Test set according to the provided algorithm: **20%**
3. Implement the kNN Algorithm: **40%**
4. Tune k according to validation set & Prepare report table according to the following format: **20%**

k	Classification Accuracy on Validation Set
3	0.8
5	0.85

5. Code in such a way so that during evaluation, new dataset can be loaded & your algorithm can be trained, validated & tested on that: **10%**

Dataset Splitting Algorithm:
Randomly Split the dataset into Training (**70%**), Validation (**15%**) and Test (**15%**) set

```
Train_set = [ ], Val_set = [ ], Test_set = [ ]
Shuffle your dataset list

for each sample S in the dataset:
    generate a random number R in the range of [0, 1]
    if 0<=R<=0.7:
        append S in Train_set
    elif 0.7<R<=0.85:
        append S in Val_set
    else:
        append S in Test_set
```

k-NN Classifier Algorithm:

K = 5

1) for each sample V in the VALIDATION set:
2) for each sample T in the TRAINING set:
3) Find Euclidean distance between Vx (features->N-1) and Tx (features->N-1)
4) Store T and the distance in list L
5) Sort L in ascending order
6) Take the first K samples
7) Take the majority class from the K samples
(this is the predicted class for sample V)
8) Now, check if this class is correct or not
9) Calculate validation accuracy =
(correct VALIDATION samples)/(total VALIDATION samples)

- Calculate validation accuracy in a similar way for K = 1, 3, 5, 10, 15, ...
- Make a table with 2 columns: K and Val_acc (report doc file)
- Now, take the K with highest Val_acc
- Use this best K to determine Test_acc (Simply replace the VALIDATION set of line 1. with TEST set)

Instructions:

- Implement in Python
- **DO NOT USE** libraries such as: "**Sklearn**", "**Scikit learning**"
- Generalize data loading and generation predictions from classifier so that you can easily run the training and evaluation on new dataset during viva