

## **#1 Classification**

For the first Classification details, we decided to use KNN with an odd number of samples(n) to guess the label values of the given test values. By using the sklearn python package, we were able to implement machine learning.

### Methodology:

1. Converted TrainData, TrainLabel, TestData files from txt to csv by removing unnecessary spaces and putting commas between each feature.
2. Import specific libraries such as pandas, numpy, and sklearn.
3. Load csv files using `pd.read_csv()`
4. Convert any rogue or NaN values to 0 such that the data could be consistent.
5. Trained the formula by using `KNeighborsClassifier` from the sklearn package then used `knn.fit()` to create the plot.
6. Used `knn.predict()` to predict each row of data in TestData.
7. Converted the result to the pandas dataframe then exported the result to a csv file.

## **#2 Missing Value Estimation**

For missing value estimation, we decided to implement the K-nearest neighbor algorithm, with a varying k value. We saved our results for missing value estimation as KimKNNDataResults.

### Methodology:

1. Converted MissingData files from txt to csv
2. Imported python libraries (pandas, numpy, and sklearn.impute KNNImputer)
3. Load csv files as pandas dataframe
4. Estimated K value by choosing odd numeric estimate of  $k=\sqrt{n}$ , where n is equal to the number of samples
5. Use `KNNImputer` from `sklearn.impute` and set nearest neighbor value as previously estimated K value from step 4
6. Convert resulting data array into pandas dataframe
7. Export resulting data frame as a CSV file