# CAP5610 HW3 - Yuan Du

Code stored at my Github: https://github.com/YuanEldaif/CAP5610

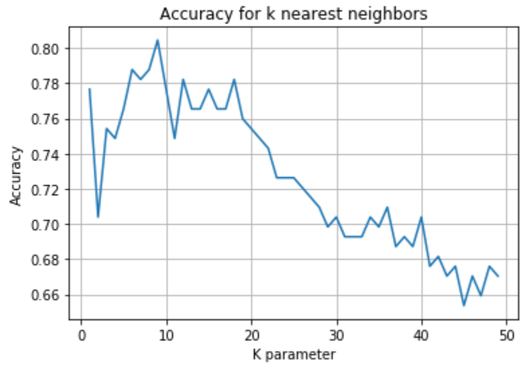
## Task 2 Programming and Evaluation on A Large Dataset (Titanic):

### Q1: Test your naïve Bayesian classification on the Titanic dataset. Report the average Accuracy, Precision, Recall, and F1 score of your five-fold cross validation. The five-folds of the Titanic data are split randomly. What do you observe and learn by applying Bayesian learning to small datasets and larger datasets?

### Q2: Implement KNN classification from scratch, and evaluate how K impacts the overall accuracy of kNN on the dataset. Plot the accuracies of kNN over k, and identify the best K. You can read sample code and try to implement by yourself. Below are some sample implementations from Github for your fast references:

```
((712, 4), (712,), (179, 4), (179,))
```

```
C:\Users\alice\anaconda3\lib\site-packages\ipykernel_launcher.py:19: TqdmDeprecationWarning: This function will be re
moved in tqdm==5.0.0
Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
```




```
The best K is at K =  9 with Maximum accuracy:- 0.8044692737430168
```