# INFO 6105 – Assignment 1

Submitted by – Parag Shah

1. The K nearest neighbor algorithm has hyper-parameters in addition to K. Use the titanic dataset to simultaneously tune the values of more than one hyper-parameter. E.g. Value of K, weight, and metric. You can either do nested for loops for this, or use the grid search OR random search function from scikit-learn. Recommendation is you submit code inside of a Jupyter notebook.

Solution: Refer **Assignment\_1\_Parag.ipynb**

1. What do you think some of the disadvantages of kNN algorithm are?

Solution: Following can be the disadvantages of kNN:

1. kNN algorithm is a lazy learner. It does not learn anything from the training data.
2. Since the model stores all the training data, it is computationally expensive.
3. It is sensitive to irrelevant attributes and imbalanced datasets.
4. Performs slow with large datasets.
5. kNN cannot deal with missing values. We need to drop that column or impute the values.
6. We need to do feature scaling (standardization and normalization) before applying KNN algorithm to any dataset. If we don't do so, KNN may generate wrong predictions.