# CS481 Data Science
## Assignment 03
**Due Date: 31st March 2020**                                    **Total Marks: 50**

- Download Balance-Scale and Flags datasets from UCI Irvine

  1. Summarize the datasets in your words. What features (attributes) are used? [**5 Points**]

  2. Using python and KNN, NaiveBayes, and C4.5 classifiers. Compare and analyse the performance using F-measure and Accuracy. Plot performance curves and discuss. Use 10 Fold Cross Validation and random train/test split(70%, 30%) [**20 Points**]

  3. Repeat using SVM classifier. Use RBF Kernel with Chi-squared distance metric: Chi-squared distance metric: $k(x_i, x_j) = e^{-\frac{1}{A} dist_{\chi^2}(x_i, x_j)}$ where $A$ is a scalar which normalises the distances. $A$ is set to the average $\chi^2$ distance between all elements of the kernel matrix. For simplicity, use only random train/test split(70%, 30%). [**25 Points**]