Classification task

The task must be completed in Python. You can use packages you prefer. Report all results in Jupyter Notebook, uploaded to your GitHub page. All steps must be visible. Make it clear and simple. Write comments, titles and explanations. The goal of this test is not to make the perfect classifier, but to assess your understanding of this group of problems. The task should take 1-2h to finish. If it takes much longer, you are doing something wrong. The deadline for this task is **April 11, 5pm**.

1. Downloading the dataset:

   Download the Spambase dataset available from the UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets/Spambase

   The Spambase data set consists of 4,601 e-mails, of which 1,813 are spam (39.4%). The data set archive contains a **processed version** of the e-mails wherein 57 real-valued features have been extracted and the spam/non-spam label has been assigned. **You should work with this processed version of the data.** The data set archive contains a description of the features extracted as well as some simple statistics over those features.

2. Building the classifier:

   Choose any classification algorithm you prefer. Build a classification model which will be able to distinguish between spam/not spam. You should perform k-fold cross-validation.

3. Evaluating the results:

   Create a table with one row per fold showing your false positive, false negative, and overall error rates, and add one final row corresponding to the average error rates across all folds. For this problem, the false positive rate is the fraction of non-spam testing examples that are misclassified as spam, the false negative rate is the fraction of spam testing examples that are misclassified as non-spam, and the overall error rate is the fraction of overall examples that are misclassified.

4. Reporting the results:

   Make a Jupyter Notebook explaining all the steps you perform. Upload the results to your GitHub page. Make sure the repository is public and submit the link to the repository here: https://www.cognitoforms.com/ISI13/ClassificationTaskReportTheResults

   Not later than April 12, 5pm PDT