

BITS F464 - Machine Learning  
I Semester 2021-22  
Assignment #1  
Weightage: 15%  
Due Date: 12<sup>th</sup> November, 2021 (FRIDAY)

---

### **Automatic Personal Email Organizer**

Build a Machine Learning model which will automatically label your emails as per the labels defined by you. You must be familiar with the labelling feature of Gmail. You have to manually create labels (or use existing labels) for your email account and then manually label your emails. You are required to automate the process of labelling. You need not consider nested labels.

#### **Data:**

Use your own emails and labels to train a model. Extract features from emails and create a dataset of reasonable size (at least 1000 emails and 6 labels). It will be a good idea to have a default label – Miscellaneous which can be used for emails which do not belong to any existing label.

#### **Machine Learning Models:**

Please note that you are not allowed to use Deep Learning Models. Use 1-gram, 2-gram, and 3-gram language models to extract features from emails. Try out different ML models and compare them.

#### **Group Information**

You can work in groups of maximum 3 students! Each group member will need to create his/her own dataset using his/her mail account and train own models. Models can be same within the group.

You can implement the same in R or Python!

#### **What you need to submit?**

- One page writeup on how you modelled the problem
- Dataset
- Code file
- Results and Comparison (Graphs & Tables)

Submit everything as a zipped file!

All the best