# TOPIC MODELLING TO CLASSIFY EMAILS IN TOPICS AND UNSUPERVISED LABELLING OF EMAILS

## PROPOSAL:

Given a data set of emails, classify them under different topics which have some kind of similarity and using this as a feature to do unsupervised labelling of emails.

## INTRODUCTION:

Emails have become the basic part of most of our lives. Our communications, information networks etc. are heavily based on emails now a days. But we receive a lot of emails per day, as a result the managing these mails becomes quite a daunting task. So we want to give automatic label to each email which is based on the content of emails.

## APPROACH:

We will use semantics of email as criteria for our topic modelling. First we will pre-process the email data to remove header, hyperlinks, numbers and other unrelated information for topic modelling. For pre-processing we can use regular expression to remove these things. Then we will give our data set as input to LDA (topic modelling algorithm).This algorithm has some parameters by which we can improve our topics. Then we use this topic as a features based on which we will cluster these mails according to their semantics. We will give a suitable label to each cluster depending on its extent of similarity with different topics.

## DATASET:

We want to implement this on our personal inbox of emails. We also can test this on online email database.

**RELATED WORK :**

There are some M.tech thesis on email clasiification using LDA.  Topic modelling has many types of application in text mining. They have similar type of use while working on data set of documents or text books.

**REFERENCES:**

- *Ozcaglar, Cagri. (2008). Classification of Email Messages Into Topics Using Latent Dirichlet Allocation:* M.S. Thesis Submitted to Rensselaer Polytechnic Institute, Troy, New York
- http://clc.yale.edu/2011/10/07/      Topic modelling Discussion
- https://code.google.com/p/topic-modeling-tool/   Topic modelling tool
- http://cs229.stanford.edu/proj2010/HarwathJohriYin-AnUnsupervisedApproachToEmailLabelSuggestions.pdf