# National University of Computer & Emerging Sciences Karachi Campus

# AI ( AI2002 ) Project Proposal

## Classifying Emails between Normal and Spam using Naive Bayers

***Section* :7A**

*Group Members:*

Syed Ahmed Ali ( 19k-1353 )
Muhammad Aashir ( 19k-0314 )
Haris Altaf ( 19k-1372 )

1. Project Motivation

● Problem: Enron was one of the biggest energy companies at the time in the US, but it collapsed in 2000 because of a significant allegation of fraud using emails. Our task would be to find out anomalies in the emails.

● Strategic Goal: To differentiate between spam/ abnormal and normal emails in order to prevent frauds.

2. Problem Definition

● Relevant factors: words, probabilistic values.

3. Relevant Method/Model

●Output: Emails to be classified among normal and abnormal from a given data set.

● Input: Enron's dataset of the emails. http://www.enron-mail.com/email/

4. Performance Measurement

● Measurement of accuracy: After training the model on the data set, we will give the model a few emails to classify which we will know that they are normal or spam. And then the results will

be compared.

- Minimum level of accuracy: 90%

5. Risks and Dependencies
    a. Constraints
    - Hardware dependencies for training model. Have a mid range GPU and CPU.
    - Data dependencies for training model.

    b. Risks
    - Hardware constraints might not be able to take on the huge data set.
        - To tackle the situation, data set size may be lowered to a size that will not hurdle the hardware and the size will be large enough to have 90% + accuracy.