# Personality Prediction using MBTI Dataset

Nalisha Rathod(gs9440)

Sujata Gorai (hc6837)

**A. Overview of the Problem**

**A.1 Literature Review**

Machine learning for predicting personalities has been a significant focus in Natural Lan- guage Processing research over the past few years. This is in part due to the availability of large corpora of online social interactions. (MBTI) Myers-Briggs Personality Type Daatset published by Kaggle competitions which have allowed researchers to gain access to datasets with 8600 training examples of post by different people around the world. In terms of methods, most research takes a text classification approach for personality prediction.

**B. Dataset Overview**

**C. Methods**

**Import the Dataset**

In each model, we have imported dataset using python. The dataset is on .csv format.

**Data Pre-processing**

We have done several methods to preprocess the comments before passing to the training model.

* Replaced "|||" from text with " " Join all texts written by 1 Person
* Removed Links from text and replace them with 'Link'
* Removed punctuation from Text