

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.
 - a. Kim Li -- kimli2 (captain)
 - b. Kevin Tzeng -- ktzeng2
 - c. Shreyas Chandrashekar -- svc3
2. What is the type of your project: Is it Data Set Creation or Leaderboard Competition Creation?
 - a. Data Set Creation
3. If your project is Data Set Creation, what is the novelty of your data set as compared with all the existing data sets? Which of the existing datasets is the closest to yours? What new task can your new data set be used to evaluate? How do you plan to create the data set?
 - a. Our dataset will be focusing on email importance and ranking them on a scale from most important to spam emails. There are a lot of methods for spam and fraud email detection which is helpful to remove unnecessary emails, but it is more beneficial for users to be able to sort through the most relevant email first, in order from most to least important. There exists email datasets already, but they are specific to a particular topic and all relevant to the user, meaning that they have already filtered out the unimportant content. Through our large range of emails, algorithms can use our data set to evaluate how good they are at analyzing the text data and highlighting key words and phrases that the email recipient may believe to be important. We plan to create the data set by combining a selection across multiple datasets, so that there is a large variety within our new dataset. We will be analyzing the number of recipients, word semantics, and other email specific attributes such as attachments or email signatures for ranking.