**END GOAL: Commercial Email Labeling**

**(April 19, 2019)**

* Finish Parsing 1 if BCC 0 otherwise, and CC which is = 1 if it has a CC but 0 otherwise (Arjun)
* Change BCC and CC to Binary (0 and 1’s) Arjun
* Represent message body and (which word embedding) <KEYWORD EXTRACTING/ SENTIMENT ANALYSIS> (June), X-cc determiner
* Day of the Week (one-hot) (Pedro)
* Process datetime (June)
* Delete message (Pedro)
* Refine Clustering Rajan
* Models (with boosting and bagging for each one and OPTIMIZE!)
  + MLPCLassifier, LSTM (Pedro)
  + KNN (Rajan)
  + DT (Arjun)
  + RF (June)
  + At the end ensemble (Everybody) Ignore this step for now

**(April 12, 2019)**

Tasks for next meeting:

* Reduce Labels (explain procedure like which ones were dropped, why?) Based on what? Explain (Rajan)
* Ignore (EVERYBODY):
  + File
  + Once folder is parsed into labels, drop.
  + X\_from
  + X\_to
  + Employee
* Parse message to have a binary column which is 1 if BCC 0 otherwise, and another column for CC which is = 1 if it has a CC but 0 otherwise (Arjun)
* Add message length as another column (Pedro)
* Preprocess (make csv file smaller) (Pedro)
* Use google model to represent message to an appropriate representation AND look to other NLP algorithms that might make this easier (June)