Choose subreddits.	Subreddits should not be too dissimilar that the model is completely uninteresting. For example, please do not choose r/finance and r/anime. The best projects choose subreddits that are similar in nature.	22 Sep
Scrape subreddits.	You should have 600 (ideally 1000+) posts per subreddit. These posts should contain text. They should also be unique (i.e. not reposts, or API returning duplicates). *Do not use metadata in your models*. **Please be neat** in order to ensure reproducibility. Beware that you may have to choose a different pair of subreddits in case you can't scrape enough unique text. Remove the names of your subreddits from both your subreddits.	22-24 Sep
EDA	You should at least understand what most of the posts are about and how your two subreddits are different. You may wish to tokenize to find what the most frequent trigrams, bigrams and unigrams are. Please remove any urls at this stage.	25 Sep
Preliminary Model Building	Please use pipelines with two different vectorization strategies (with/without) and two different models. By this time you'll already know logistic regression and naive Bayes. If your model can distinguish between your subreddits with zero difficulty, you should probably choose more similar subreddits.	26 Sep
More Model Building	Use the models that will only be taught in week 6, such as RandomForest, support vector machines.	27-30 Sep
Conclusions	Compare and contrast your models on the subreddits you chose. Why do you think each model performed well (or why not)?	Continuous; you should always be assessing your models critically with a view to building understanding
Prepare slides and rehearse your presentations.	Come up with slides to tell us what you did. 8 minutes each, so please make a maximum of 10 high-impact slides.	30 Sep