Which tasks have been completed?

I have determined the three techniques I wanted to implement (naive bayes, knn, and max entropy classifiers).

I have generated my datasets, and have generated the feature vector representation of each URL. I'm still experimenting with different representations but have most of the code ready for me to easily make changes to it. I originally tried using tokens from the URL but have moved to using English words in the URL.

I have also created the first draft of my naive bayes and knn classifiers. They both currently classify about 80% of URLs (n=500) correctly.

Which tasks are pending?

I need to implement the max entropy classifier and tune my models.

I also need to prepare the documentation and potentially add some files to be ran specifically for demoing.

Are you facing any challenges?

Progress has been limited at times because of hardware limitations (I considered moving to cloud but wanted to keep it simple/local for demo purposes). So I haven't been able to use my entire dataset yet which may skew the classifiers performance.