**CS434 Proposal DRAFT**

LINKS THAT HELP:

<https://www.kaggle.com/fernandocanteruccio/quoras-question-pairs-modeling-notebook/comments/notebook>

<https://www.kaggle.com/c/quora-question-pairs/discussion/30411>

pretrained models:

<https://www.kaggle.com/c/quora-question-pairs/discussion/30286>

<https://www.kaggle.com/c/quora-question-pairs/discussion/30340>

https://aclweb.org/anthology/K15-1013

As the second step of the final project, you need to research your chosen competition and submit a brief proposal. Note that both the [Data Science Bowl](https://www.kaggle.com/c/data-science-bowl-2017" \t "_blank) and the [Two Sigma Connect challenge](https://www.kaggle.com/c/two-sigma-connect-rental-listing-inquiries" \t "_blank) have closed, but you can still submit your predictions and get evaluated. The link that I sent out previously may not work anymore.

Your proposal should include:

* Your team composition
  + Michael Lee, Alex Nguyen
* Which competition you will participate in
  + Quora Question Pairs
  + https://www.kaggle.com/c/quora-question-pairs
* Your plan of attack – what approaches, how will you div up your work
  + Method:
    - Python
    - Keras as a platform
  + Data prep:
    - Pandas
  + Data modeling:
    - Decision tree over the words and find the pair with most similar pairs of words?
    - CNN(https://aclweb.org/anthology/K15-1013)
* Questions you may have
  + Are we competing against each other in the competition?

To prepare for the proposal, **you need to do some research**

* Read up the discussions on the discussion forum of the competition. It contains a lot of information about preprocessing, representation of the data, and approaches people have tried
* The closed competitions have discussions on the approaches taken by the leader – read up. You don’t have to reinvent the wheel, you can build on successful approaches.