DS-GA 1006 - Capstone
Team Formation and Project Selection

1. Team name:
    a. Beyond Google -- if assigned to word embeddings project
    b. Cancer Detection Squad (CDS) -- if assigned to cancer project
    c. Helix Hackers -- if assigned to DNA project
2. Team members:
    a. Daniel Amaranto - 2nd year DS student. Econ undergrad with research experience that includes surgical outcomes, cancer patients, industrial parks, franchise development, and machine learning methods in criminal prosecution decisions. Looking to advance predictive analytical capabilities by learning new methods and applying them to different data entities (pictures, text, audio, etc.)
    b. Brenton Arnaboldi - second-year student at CDS interested in Natural Language Processing (NLP) and machine learning. Prior projects include using random forests to predict the next pitch from an MLB pitcher and applying Doc2Vec to map judges' text opinions into "belief vectors". Over the summer, he worked at a pharmaceutical company (Axovant Sciences) to identify outliers from clinical trials for Alzheimer's Disease. He studied Math and Economics as an undergrad at Amherst.
    c. Akash Kadel - 2nd year DS student too. Prior to joining NYU CDS, I completed my Bachelors in Computer Science and interned at two Companies as Data analyst. I have experience in Software development, competitive programming, visualization and data modeling. In my last summer (2017) I interned for Amazon Kindle team and mainly worked with Java and Spark for building a new software for them. Looking forward to extend my Data science knowledge in a much more practical academic project.
3. Reason for teaming up.

    We were placed in a larger cluster of students (Group 3) during the second capstone class, and the three of us decided to form a team because of our shared interest in social science and medical issues. We have each taken the core CDS classes up to this point (including Machine Learning), but have not taken Deep Learning (although we had a brief introduction to neural nets in Machine Learning). We are extremely looking forward to tackle problems that would give us a good introduction into optimization with neural nets.

4. Top 3 project choices

    1) Princeton University - Word embeddings for quotes

    This project is unique and very interesting to each of us. It would allow us to study a system that we are already familiar with and sometimes use (Google's smart reply), but more importantly it would give us an opportunity to improve it in a novel way. Between us we have some experience with basic NLP and word/document embeddings (e.g. Doc2Vec), and we are all currently taking NLP with Professor Cho; we will likely be studying methods that can be directly applied to this work. Moreover, deploying our work into a Chrome extension would be a great way to showcase the system we develop. Akash has a background in Java and HTML, abilities specified in the project proposal.

2) Breast Cancer Detection with Recurrent Attention Model

This project appeals to us because we have the requisite skill set in machine learning and quantitative analysis, some basic medical knowledge, and a familiarity with deep learning methods that we would like to hone. Dr. Cho's other similar projects, Interpretable Model for Breast Cancer Detection and Augmenting breast cancer screening classification models with auxiliary data resources, are also of interest to us.

3) Vector institute for Artificial Intelligence - Decoding the regulatory genome

We like this project because the problem is interesting and well-defined. Furthermore, as indicated earlier, we are intrigued by projects on medical issues, and a few members of our team took Biology courses at the undergraduate level. While we don't have extensive experience training deep neural networks, we understand the core concepts behind neural nets based on our coursework in Machine Learning and NLP, and we'd like to think we are "quick learners".