COVID-19 Research:

Meeting 1( Friday 4/10/2020):

1. Scope of the project.
2. Identified medical care as the task.
3. Exploring interventions as the subtask.

Work done:

1. Looked at pubmed data and kaggle data.
2. Extracted all json files of articles and created a unified dataset/table.
3. Created trivial features such as word count, character count, average word length, number of stopwords, number of special characters, number of numerics.
4. Cleaned the body text of every article(Removed stop words, lowercased words, removed punctuations and special characters).
5. Looked at most frequent words and rare words in the overall documents.

Potential ways of going about:

1. Filter and work on subtask only
2. Create models such as bag of words model and term frequency inverse document frequency model, word2vec models(skipgram, continuous bow) or glove models.(or use bert)
3. Search for top 10 topics in every document.

Meeting 2(Tuesday 4/14/2020):