**Prompt Learning – Health Prompt**

Week 1-3 (Jan 3, 2022 – Jan 9, 2022):

* Went through many papers on prompt learning and found out that AutoPrompt by UC Irvine(EMNLP 2020 : <https://github.com/ucinlp/autoprompt> ) and OpenPrompt by Tsinghua University (<https://github.com/thunlp/OpenPrompt>) are the available libraries for prompt learning.
* These are in their initial stages and still lacking many features.
* successfully got both the libraries working on general data, but open prompt seems more relevant to our project as it is a zero-shot learning methodology, while autoprompt still needs training dataset for selecting the best prompt template. The caveat is that openprompt needs manual templates defined by us, hence the best template can only be found by trial and error.
* tried Openprompt with health datasets and it gave good results on Clinical BERT and BERT. BioBERT and RoBERTa showed poor performance for some reasons. But even BERT failed to classify covid related texts.
* Results from larger text corpora(i2b2) was not promising. I can demonstrate that in the meeting tomorrow.
* Used streamlit to deploy the prompt application on university server and an external server. This will be useful when we are validating the results with a medical expert.
* I have seen papers describing that prompt can be used for NLU, NLP and other NLP tasks. Probably, we can try some of these once we are in good shape.

Week 4 (Jan 10, 2022 – Jan 16, 2022):

* Try with Smoking Dataset
* Try 3 BERT Model
* Try weighted score for large text corpus
* Solve negation issue
* Read both papers and solve the template issue
* Make autoprompt working
* Create a package out of this(can be pushed to next week)