SODNER:

1. Separate the text and ann files in different folders.
2. Use the truncated text and ann as BERT can only take 512 tokens.
3. Also make changes to the models as per the allennlp checkpoint.
4. Run dai et al and preprocess the data and get train.txt, dev.txt and test.txt.
5. Pass these .txt files to SODNER preprocess code and get train.json, dev.json and test.json.
6. Downgrade python to 3.8 and install all dependencies.
7. RUN SODNER model on google colab on these json files.
8. Save the model.tar.gz
9. **IMPORTANT: Unarchive model.tar.gz (**tar -xvf model.tar.gz) **and change the path of the pretrained BERT model in weights.tar.gz and vocab path in config file to original pre trained model.**
10. **Archive it again by the name of model.tar.gz using python code in Pycharm.**
11. **Make changes in predict.py file in allennlp library to get the predictions in desired format.**
12. **Run bellow command to get the predictions.**

allennlp predict /Users/shashankgupta/Documents/Pycharm/S0dn3r/models/Raredis\_0001/model.tar.gz /Users/shashankgupta/Documents/Raredis/SODNER/corrected\_ano\_preprocessfile/truncated\_doc/test.json --include-package sodner --predictor my\_predictor --output-file /Users/shashankgupta/Documents/Raredis/SODNER/corrected\_ano\_preprocessfile/truncated\_doc/prediction3.json