CB&B 7400 Homework 3

Installation of PyDeid

Please install PyDeid [1, 2] from the following GitHub repo:

https://github.com/GEMINI-Medicine/pyDeid

Here are the code in Colab to install PyDeid and its dependency:

!pip install git+https://github.com/GEMINI-Medicine/pyDeid

!pip install Faker

After installation, please run the example shown at the end of the following section:

https://github.com/GEMINI-Medicine/pyDeid?tab=readme-ov-file#2-local-installation-and-setup

De-identifying Clinical Texts

Please test PyDeid using the following 3 clinical tests:

1. https://dss.mo.gov/mhd/cs/psych/pdf/progressnote_indv_sample.pdf

For this one, please use the text starting from "Recipient Information" and ending at "ground her further." (both inclusive). If needed, please remove all "new line" characters before running PyDeid. After running it, please use a word processing tool (e.g., Microsoft Word or Google Doc) to annotate the "de-identified" clinical text (*not* the original one). Please mark the changed parts in yellow. Please note that first name and last name should be marked separately. Also, please mark the unchanged parts that you feel like should have been changed in green. Below is an example using the sample text from the PyDeid GitHub page mentioned above:

Jared Gentry starred in The Lord of the Rings, released on 1975. 16 51975-01-06.

2. https://svn.apache.org/repos/asf/ctakes/branches/ctakes-3.2.3/ctakes-clinical-pipeline/src/test/data/plaintext/testpatient_plaintext_1.txt

Similarly, please de-identify and mark the results in the same file above.

3. https://svn.apache.org/repos/asf/ctakes/branches/ctakes-3.2.3/ctakes-clinical-pipeline/src/test/data/plaintext/testpatient_plaintext_2.txt

Please do the same for this clinical text.

Submission

Submit your results in a single PDF (including the results of 3 clinical texts) by 11:59pm on Tuesday 9/30.

Grading

20 points total, each “yellow” marking count for 1 point, and the rest of the points are given when the PyDeid outputs are correct. The “green” marking is optional and will not count towards the scores.

References

1. https://doi.org/10.1093/jamiaopen/ooae152
2. https://doi.org/10.1186/1472-6947-8-32