

The goal of this task was to run the Community Notes (CN) algorithm from Nov 2023 with data from Nov 2023.

To accomplish this, I cloned the CN repo from the latest commit in October 2023 (After that date in October, the next update was after Nov 1st, so we can assume the code from the last update in October was being used in Nov 1st).

Then I downloaded all of the data files and used Python and Pandas to filter the rows with a timestamp before or equal to Nov 1st, 2023. This is what my notebook file in the Pull request shows. I then copied the filtered data files to my cloned CN repo. I tried to run the CN algorithm however, since the code is a couple of years old, the algorithm expected a different set of columns than were present in the data files. By looking at the CN code, I determined which columns I needed to delete from my data files. I used Python and Pandas to programmatically drop the columns.

Additionally, there are multiple ratings available on the download page. I wrote a quick Python function to filter each ratings file by rows with a timestamp before or on Nov 1st 2023, and also drop the extraneous columns. Then I concated all of the ratings files into one to feed into the CN algorithm. The resulting ratings file was 3.1 Gb.

Finally, I was able to get the CN algorithm to run with the filtered files. At the time of this writing, the algorithm has been running for 12+ hours.

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS BRUNO LOGS
python3.9 - sourcecode

TRAIN FIT LOSS: 0.07425021380186081
Global Intercept: 0.1297459602355957

Users: 212803, Notes: 241372
initializing notes
initializing users
initialized global intercept
cpu
epoch 0 0.2517623007297516
TRAIN FIT LOSS: 0.21868611872196198
epoch 20 0.12439025938510895
TRAIN FIT LOSS: 0.09599754214286804
epoch 40 0.11511796712875366
TRAIN FIT LOSS: 0.09011238813400269
epoch 60 0.11376404762268066
TRAIN FIT LOSS: 0.08862495422363281
epoch 80 0.11359762400388718
TRAIN FIT LOSS: 0.0884496197104454
epoch 100 0.1135750412940979
TRAIN FIT LOSS: 0.08846316486597061
Num epochs: 106
epoch 106 0.11357376724481583
TRAIN FIT LOSS: 0.08845935761928558
Global Intercept: 0.16019900143146515
```