1. I would create a .gitignore and ignore the /data path.
   1. This is because we can’t upload very large documents to github, our data will be multiple GB’s
2. Inside the data folder, create folder named CodeSearchNet
   1. Get the zip files from here <https://huggingface.co/datasets/code-search-net/code_search_net/tree/main/data>
   2. Place the code language you want to use in.
   3. You can alternative use “datasets.load\_dataset()” to get the data but will get all languages.

A screenshot of a computer

Description automatically generated