**LLM Benchmark Projects**

**06 Sept, 2024**

1. Data C++: [PR2Commits\_perf\_issue\_C++ - Google Sheets](https://docs.google.com/spreadsheets/d/1PubFVYm6KShQ81B8ZqJq4lTmJ79djYhmYT5MehZFhF8/edit?gid=0#gid=0)
2. **Task-01:** Design a scrapper to separate codes
3. Pull request (PR) demonstrates a proposed code change
4. [JSON Lines](https://jsonlines.org/)
5. Scrap lib: [beautifulsoup4 · PyPI](https://pypi.org/project/beautifulsoup4/)

**07 Sept, 2024**

1. Generated 300 entries in jsonl file but need to check it with Prof if the format is desired

**09 Sept, 2024**

1. Meeting minutes: This is a new dataset: <https://docs.google.com/spreadsheets/d/1qDVTIzfJw70eB6KU61GV_JBFrMilPelAXaakvZz4350/edit?gid=0#gid=0>
2. Task-01: Need to verify this dataset if all the entries contain only 01 function change
3. Task-02: Design a scrapper on this dataset. Output entries: Commit title, Code difference (+,-), Before version, after version