Reproducing and replicating results

You have chosen 1 paper, and are doing an effort to reproduce and replicate results, relying on tools and methods of prior lab sessions.

The work to be done is a git repository with the code, data, and any artefact to reproduce and replicate results as well as a README.md structured as follows:

- A short *introduction* paragraph briefly presenting the subject and the problem, as well as key results/insights you obtained
- A *reproducibility* section
    - a subsection that describes *how to reproduce results* from a technical and software point of views
        - it is expected to propose a reproducible environment (through Dockerfile) as well as automated scripts to produce data and analyze data (the former part can be done through Jupyer notebooks and tools like nbconvert https://nbconvert.readthedocs.io/en/latest/)
        - bonus: develop a Github action to produce and/or analyze data
    - a subsection that reports any encountered issue or deviation compared to the original study, including a description of what you improve (if any)
    - a subsection that answers the question "*is the original study reproducible?*" based on the reproduction and presentation of results
- A *replicability* section
    - a subsection that lists all *variability factors* that could be considered
        - for each variability factor, include the possible values that can be taken, as well as a description of why it is relevant to consider this factor as part of a replicability inquiry.
        - document constraints across variability factors (if any)
        - bonus: develop a command-line interface (CLI) that can be used to explore all variability factors (and their combinations) through CLI parameters[1]
    - a subsection that explains one or several replications that you execute including
        - instructions on how to execute your replication
        - presentation and analysis of results
    - a subsection that answers the question *"does it confirm the results and conclusion of the original study?"* based on the replication(s) and presentation of results
- A *conclusion* section that wraps up your findings and opens perspectives on limitations and future work to be done

---

[1] This "bonus" does not aim to implement all variability factors, the CLI serving as a way to define the variability space, and it's up to others to fully realize all (combinations of) parameters. As part of your replication effort, you will most likely implement a subset of variability factors

Deadline: 18 December 2024 (23:59 Paris time)

A README.md template is provided

Project Title

Introduction

Briefly introduce the subject of the study, the problem it addresses, and the key results or insights obtained from the reproduction and replication effort.

Reproducibility

How to Reproduce the Results
1. **Requirements**
   - List dependencies and their versions (e.g., Python, R, libraries, etc.).
   - Specify any system requirements.

2. **Setting Up the Environment**
   - Provide instructions for using the Dockerfile to create a reproducible environment:
     ```bash
     docker build -t reproducible-project .
     docker run -it reproducible-project
     ```

3. **Reproducing Results**
   - Describe how to run the automated scripts or notebooks to reproduce data and analyze results:
     ```bash
     bash scripts/run_analysis.sh
     ```
   - Mention Jupyter notebooks (if applicable):
     Open `notebooks/reproduce_results.ipynb` to execute the analysis step-by-step.

4. **Automation (Bonus)**
   - Explain the included GitHub Action that produces or analyzes data automatically.

Encountered Issues and Improvements
- Report any challenges, errors, or deviations from the original study.
- Describe how these issues were resolved or improved, if applicable.

Is the Original Study Reproducible?
- Summarize the success or failure of reproducing the study.
- Include supporting evidence, such as comparison tables, plots, or metrics.

Replicability

Variability Factors

- **List of Factors**: Identify all potential sources of variability (e.g., dataset splits, random seeds, hardware).

  Example table:

[2] | Variability Factor | Possible Values | Relevance |
  |-------------------|--------------------|---|
  | Random Seed | [0, 42, 123] | Impacts consistency of random processes |
  | Hardware | CPU, GPU (NVIDIA) | May affect computation time and results |
  | Dataset Version | v1.0, v1.1 | Ensures comparability across experiments |

- **Constraints Across Factors**:
  - Document any constraints or interdependencies among variability factors.
    For example:
    - Random Seed must align with dataset splits for consistent results.
    - Hardware constraints may limit the choice of GPU-based factors.

- **Exploring Variability Factors via CLI (Bonus)**
  - Provide instructions to use the command-line interface (CLI) to explore variability factors and their combinations:
    ```bash
    python explore_variability.py --random-seed 42 --hardware GPU --dataset-version v1.1
    ```
  - Describe the functionality and parameters of the CLI:
    - `--random-seed`: Specify the random seed to use.
    - `--hardware`: Choose between CPU or GPU.
    - `--dataset-version`: Select the dataset version.

Replication Execution
1. **Instructions**
   - Provide detailed steps or commands for running the replication(s):
     ```bash
     bash scripts/replicate_experiment.sh
     ```


2. **Presentation and Analysis of Results**
   - Include results in text, tables, or figures.
   - Analyze and compare with the original study's findings.

Does It Confirm the Original Study?
- Summarize the extent to which the replication supports the original study's conclusions.
- Highlight similarities and differences, if any.

Conclusion
- Recap findings from the reproducibility and replicability sections.
- Discuss limitations of your

---

[2] It's really an example. Some factors presented here may not make any sense in your study!