AI/ML for Climate Workshop

International Livestock Research Institute (ILRI)

hide: - toc

🧪 ML Workflow for Weather & Climate (Project Structure)

- **Cookiecutter**
- **Git**
- **VS Code**
- **GitHub**

A logical, reasonably standardized but flexible **project structure** for doing and sharing data science work.

Why this matters

A well-structured project isn't just about tidiness it **saves time**, **reduces errors**, and lets you focus on solving hard problems.

🚀 Interactive Learning

launch binder

Click the Binder button above to launch an interactive Jupyter notebook for NumPy and Pandas climate data analysis!

Prerequisites

- Python 3.10+

- pip / pipx
- Git
- VS Code (or PyCharm/Spyder)

---

0) Quick Checks

```
import shutil, subprocess, sys
def which(x):
    return shutil.which(x) or "<not found>"
print("Python:", sys.version.split()[0], "| exe:", sys.executable)
print("pip:", which("pip"))
print("git:", which("git"))
print("code (VS Code):", which("code"))
print("cookiecutter:", which("cookiecutter"))
```

Output:

```
Python: 3.13.8 | exe: c:\Users\yonas\Documents\ICPAC\python-ml-gha-venv\Scripts\python.
pip: c:\Users\yonas\Documents\ICPAC\python-ml-gha-venv\Scripts\pip.EXE
git: C:\Program Files\Git\cmd\git.EXE
code (VS Code): C:\Users\yonas\AppData\Local\Programs\Microsoft VS Code\bin\code.CMD
cookiecutter: c:\Users\yonas\Documents\ICPAC\python-ml-gha-venv\Scripts\cookiecutter.EX
```

---

1) Installing Cookiecutter

Option A: pip

```
python -m pip install --upgrade pip
pip install cookiecutter
```

Verify:

```
cookiecutter --version
```

---

2) Creating Your ML Project Structure (Cookiecutter)

From the parent directory where you want your project

```
cookiecutter https://github.com/drivendataorg/cookiecutter-data-science -c v1
```

Interactive prompts (examples):

```
project_name [My Awesome Project]: MyMLProject
repo_name [my_ml_project]:
author_name [Your Name]: Jane Doe
description [A short description of the project.]: A cool ML project using Cookiecutter
```

Non-interactive:

```
cookiecutter gh:drivendata/cookiecutter-data-science -c v2 --no-input   project_name="M
```

---

3) Understanding the Structure

```
MyMLProject/
├── data/
│   ├── external/
│   ├── interim/
│   ├── processed/
│   └── raw/
├── docs/
├── models/
├── notebooks/
├── references/
├── reports/
│   └── figures/
└── src/
    ├── data/
    ├── features/
    ├── models/
    └── visualization/
```

- **data/**: external/raw/interim/processed for data lifecycle
- **docs/**: project documentation
- **models/**: saved models/checkpoints
- **notebooks/**: EDA & experiments
- **references/**: papers/links
- **reports/** + **figures/**: generated outputs
- **src/**: importable code modules

---

4) Initialize Git & Push to GitHub

From your project directory:

Initialize Git

```
git init
git add -A
git commit -m "Initial commit: Created project structure"
```

Create a Repository on GitHub

- Log in to GitHub and create a new repository.
- Click on New (or use this link) to create a new repository.
- Name it the same as your local project (e.g., MyMLProject).
- Do not initialize with a README (since you already have one or can add it later).

5) Pushing Your Project to GitHub

- Link your local repository to the GitHub repository and push your changes.
- Add the Remote Repository Replace and with your details:

```
git remote add origin https://github.com/<your-github-username>/<repository-name>.git
```

Example:

git remote add origin https://github.com/YonSci/MyMLProject.git

Push Your Code Push your initial commit to GitHub:

```
git push origin master
or
git branch -M main  # Ensure your main branch is named 'main'
git push -u origin main
```

After running these commands, refresh your GitHub repository page, and you should see your project files online!

Optional: Git Comands

```
# To check the status
git status
git status -s

# To check the commit history
git log

# To add the file to the staging area
git add -A
git add .
```

To commit the changes

```
git commit -m "First commit"

# Push the changes from the local to Github
git push origin master

# When a local branch is behind the remote branch:
git fetch origin
git merge origin/master
git push origin master
```

Conclusion

Congratulations!

You now have a well-organized machine learning project structure generated by Cookiecutter or the Pure-Python Git, VS Code, and GitHub.
