

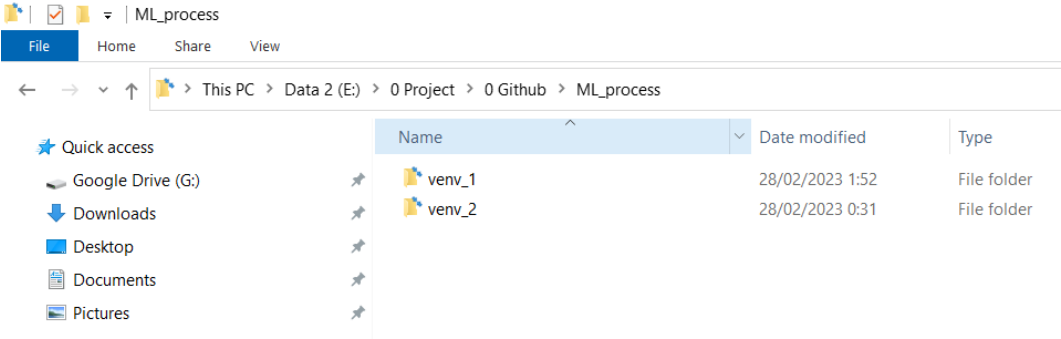
# IMachine Learning Process

## Exercise 1

Name: Jendra Riyan Dwiputra

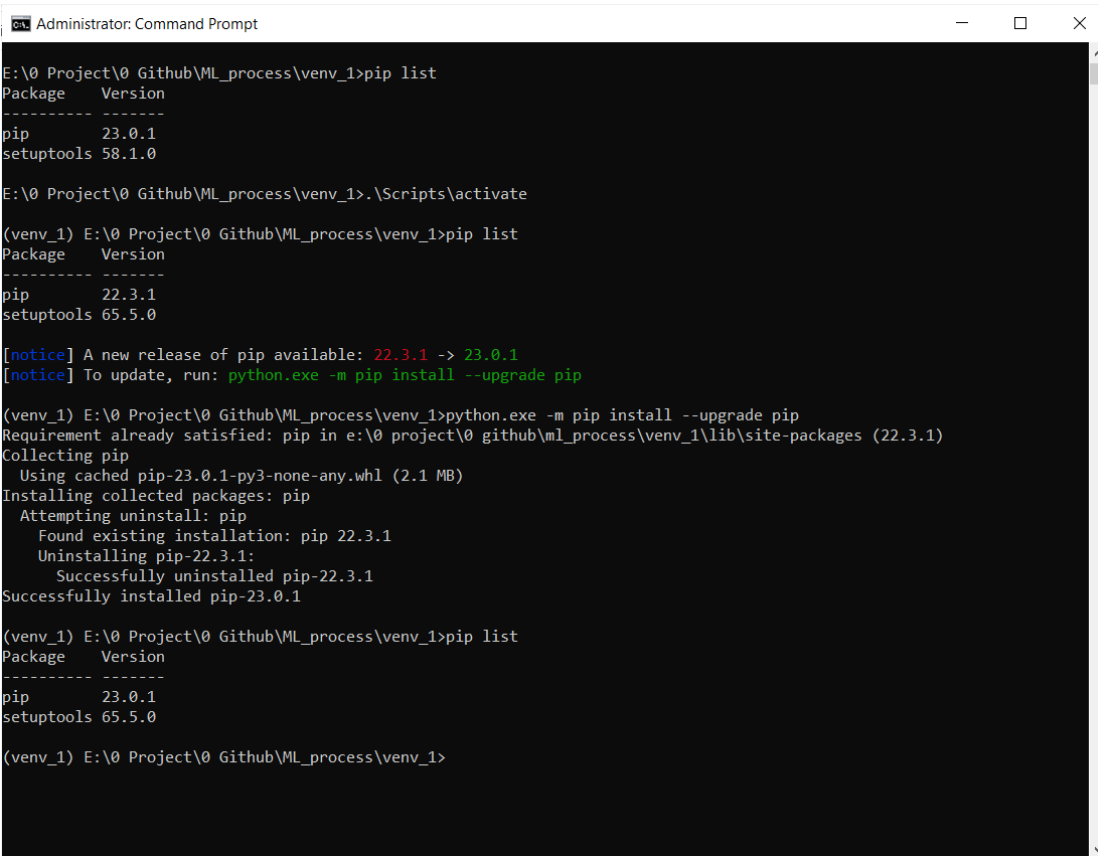
Github Link: [https://github.com/jendra/ml\\_process](https://github.com/jendra/ml_process)

### 1. Setup your local environment, activate and update your pip!



The screenshot shows a Windows File Explorer window titled 'ML\_process'. The address bar indicates the path: 'This PC > Data 2 (E:) > 0 Project > 0 Github > ML\_process'. The left sidebar shows 'Quick access' with links to Google Drive (G:), Downloads, Desktop, Documents, and Pictures. The main pane displays a table of files and folders:

Name	Date modified	Type
venv_1	28/02/2023 1:52	File folder
venv_2	28/02/2023 0:31	File folder

The screenshot shows a Windows Command Prompt window titled 'Administrator: Command Prompt'. The command prompt displays the following commands and output:

```
E:\0 Project\0 Github\ML_process\venv_1>pip list
Package      Version
-----
pip          23.0.1
setuptools   58.1.0

E:\0 Project\0 Github\ML_process\venv_1>.\.Scripts\activate

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>pip list
Package      Version
-----
pip          22.3.1
setuptools   65.5.0

[notice] A new release of pip available: 22.3.1 -> 23.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>python.exe -m pip install --upgrade pip
Requirement already satisfied: pip in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (22.3.1)
Collecting pip
  Using cached pip-23.0.1-py3-none-any.whl (2.1 MB)
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 22.3.1
    Uninstalling pip-22.3.1:
      Successfully uninstalled pip-22.3.1
  Successfully installed pip-23.0.1

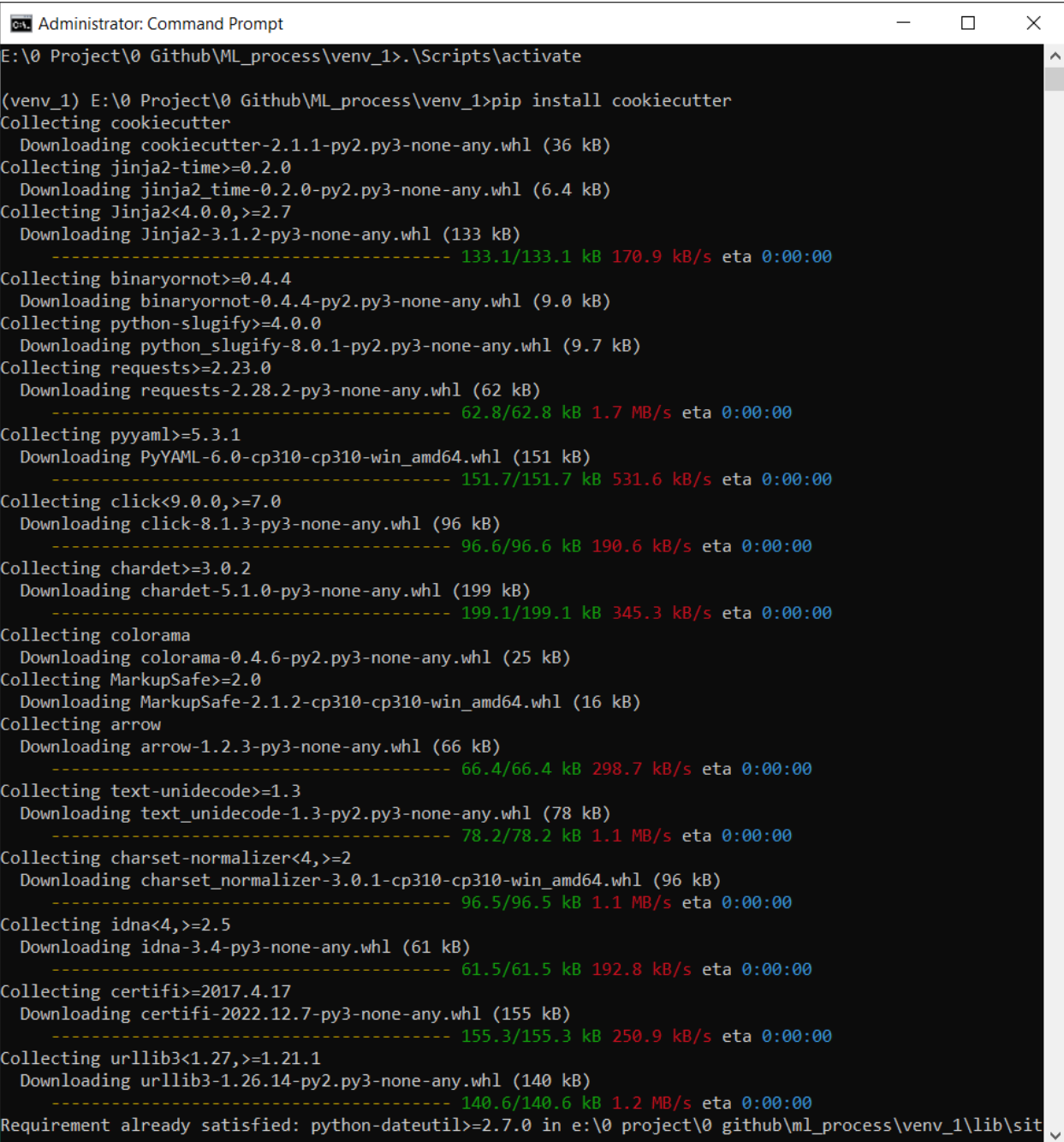
(venv_1) E:\0 Project\0 Github\ML_process\venv_1>pip list
Package      Version
-----
pip          23.0.1
setuptools   65.5.0

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>
```

## 2. Setup folder structure and exclude your venv folder

I setup using cookiecutter, based on guideline: <https://drivendata.github.io/cookiecutter-data-science/> (Asistensi 1)

- **Step 1:**  
**Activate and Install Cookiecutter**



```
Administrator: Command Prompt
E:\0 Project\0 Github\ML_process\venv_1>.Scripts\activate

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>pip install cookiecutter
Collecting cookiecutter
  Downloading cookiecutter-2.1.1-py2.py3-none-any.whl (36 kB)
Collecting jinja2-time>=0.2.0
  Downloading jinja2_time-0.2.0-py2.py3-none-any.whl (6.4 kB)
Collecting Jinja2<4.0.0,>=2.7
  Downloading Jinja2-3.1.2-py3-none-any.whl (133 kB)
----- 133.1/133.1 kB 170.9 kB/s eta 0:00:00
Collecting binaryornot>=0.4.4
  Downloading binaryornot-0.4.4-py2.py3-none-any.whl (9.0 kB)
Collecting python-slugify>=4.0.0
  Downloading python_slugify-8.0.1-py2.py3-none-any.whl (9.7 kB)
Collecting requests>=2.23.0
  Downloading requests-2.28.2-py3-none-any.whl (62 kB)
----- 62.8/62.8 kB 1.7 MB/s eta 0:00:00
Collecting pyyaml>=5.3.1
  Downloading PyYAML-6.0-cp310-cp310-win_amd64.whl (151 kB)
----- 151.7/151.7 kB 531.6 kB/s eta 0:00:00
Collecting click<9.0.0,>=7.0
  Downloading click-8.1.3-py3-none-any.whl (96 kB)
----- 96.6/96.6 kB 190.6 kB/s eta 0:00:00
Collecting chardet>=3.0.2
  Downloading chardet-5.1.0-py3-none-any.whl (199 kB)
----- 199.1/199.1 kB 345.3 kB/s eta 0:00:00
Collecting colorama
  Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Collecting MarkupSafe>=2.0
  Downloading MarkupSafe-2.1.2-cp310-cp310-win_amd64.whl (16 kB)
Collecting arrow
  Downloading arrow-1.2.3-py3-none-any.whl (66 kB)
----- 66.4/66.4 kB 298.7 kB/s eta 0:00:00
Collecting text-unidecode>=1.3
  Downloading text_unidecode-1.3-py2.py3-none-any.whl (78 kB)
----- 78.2/78.2 kB 1.1 MB/s eta 0:00:00
Collecting charset-normalizer<4,>=2
  Downloading charset_normalizer-3.0.1-cp310-cp310-win_amd64.whl (96 kB)
----- 96.5/96.5 kB 1.1 MB/s eta 0:00:00
Collecting idna<4,>=2.5
  Downloading idna-3.4-py3-none-any.whl (61 kB)
----- 61.5/61.5 kB 192.8 kB/s eta 0:00:00
Collecting certifi>=2017.4.17
  Downloading certifi-2022.12.7-py3-none-any.whl (155 kB)
----- 155.3/155.3 kB 250.9 kB/s eta 0:00:00
Collecting urllib3<1.27,>=1.21.1
  Downloading urllib3-1.26.14-py2.py3-none-any.whl (140 kB)
----- 140.6/140.6 kB 1.2 MB/s eta 0:00:00
Requirement already satisfied: python-dateutil>=2.7.0 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from cookiecutter)
Requirement already satisfied: six>=1.5 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: backports.zoneinfo>=0.2.1 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: tzdata>=2022.1 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: typing-extensions>=3.7.4.3 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: importlib-resources>=5.9.0 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: zipp>=3.8.0 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from importlib-resources>=5.9.0->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: pytz>=2022.7 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from tzdata>=2022.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: backports.zoneinfo in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from tzdata>=2022.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: typing-extensions in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from backports.zoneinfo>=0.2.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: importlib-resources in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from backports.zoneinfo>=0.2.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: zipp in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from importlib-resources>=5.9.0->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: pytz in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from pytz>=2022.7->tzdata>=2022.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: backports.zoneinfo in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from pytz>=2022.7->tzdata>=2022.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: typing-extensions in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from backports.zoneinfo>=0.2.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: importlib-resources in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from backports.zoneinfo>=0.2.1->python-dateutil>=2.7.0->cookiecutter)
Requirement already satisfied: zipp in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from importlib-resources>=5.9.0->python-dateutil>=2.7.0->cookiecutter)
```

```
Requirement already satisfied: python-dateutil>=2.7.0 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from arrow->jinja2-time>=0.2.0->cookiecutter) (2.8.2)
Requirement already satisfied: six>=1.5 in e:\0 project\0 github\ml_process\venv_1\lib\site-packages (from python-dateutil>=2.7.0->arrow->jinja2-time>=0.2.0->cookiecutter) (1.16.0)
Installing collected packages: text-unidecode, charset-normalizer, urllib3, pyyaml, python-slugify, MarkupSafe, idna, colorama, chardet, certifi, requests, Jinja2, click, binaryornot, arrow, jinja2-time, cookiecutter
Successfully installed Jinja2-3.1.2 MarkupSafe-2.1.2 arrow-1.2.3 binaryornot-0.4.4 certifi-2022.12.7 chardet-5.1.0 charset-normalizer-3.0.1 click-8.1.3 colorama-0.4.6 cookiecutter-2.1.1 idna-3.4 jinja2-time-0.2.0 python-slugify-8.0.1 pyyaml-6.0 requests-2.28.2 text-unidecode-1.3 urllib3-1.26.14

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>
```

- **Step 2:**

**Starting a new project with command:**

cookiecutter <https://github.com/drivendata/cookiecutter-data-science>

```
Administrator: Command Prompt

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>cookiecutter https://github.com/drivendata/cookiecutter-data-science
You've downloaded C:\Users\jendra\.cookiecutters\cookiecutter-data-science before. Is it okay to delete and re-download it? [yes]: yes
project_name [project_name]: ML Process
repo_name [ml_process]: ml_process
author_name [Your name (or your organization/company/team)]: Jendra Riyan Dwiputra
description [A short description of the project.]: Simple project about Machine Learning Process.
Select open_source_license:
1 - MIT
2 - BSD-3-Clause
3 - No license file
Choose from 1, 2, 3 [1]: 3
s3_bucket [[OPTIONAL] your-bucket-for-syncing-data (do not include 's3://')]:
aws_profile [default]:
Select python_interpreter:
1 - python3
2 - python
Choose from 1, 2 [1]: 2

=====
*** DEPRECATION WARNING ***

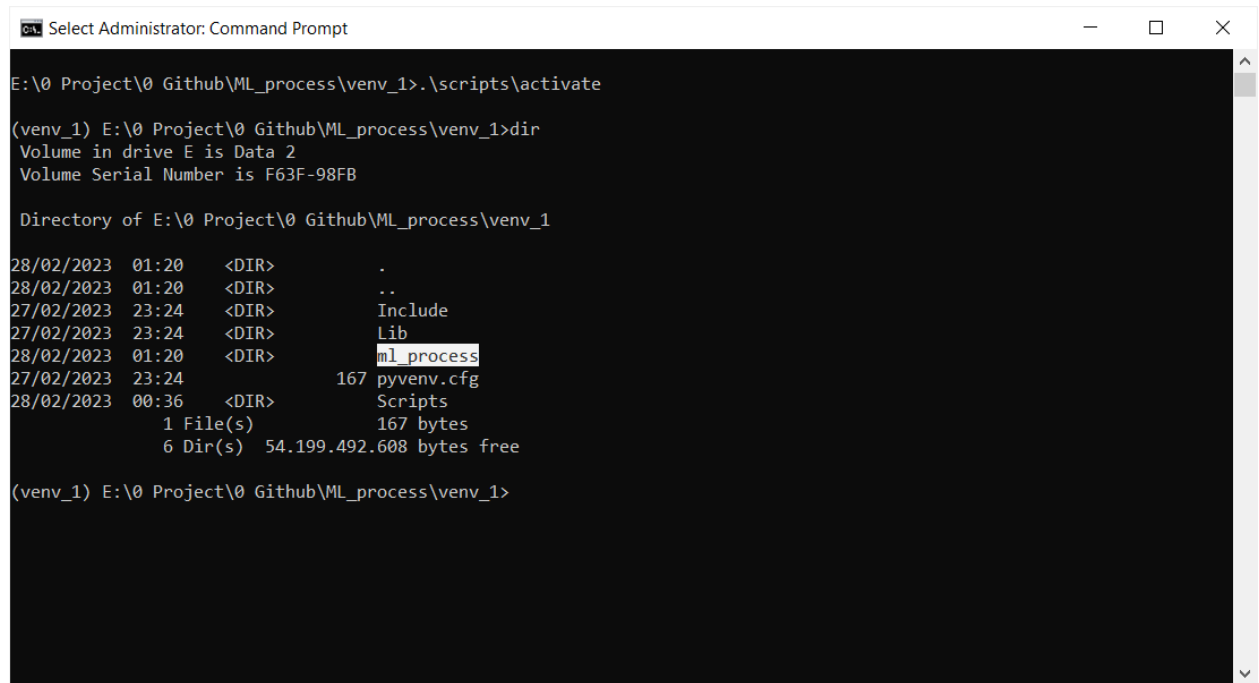
Cookiecutter data science is moving to v2 soon, which will entail using
the command `ccds ...` rather than `cookiecutter ...`. The cookiecutter command
will continue to work, and this version of the template will still be available.
To use the legacy template, you will need to explicitly use `-c v1` to select it.

Please update any scripts/automation you have to append the `-c v1` option,
which is available now.

For example:
  cookiecutter -c v1 https://github.com/drivendata/cookiecutter-data-science
=====

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>
```

- **Step 3:**  
**Check Directory**



```
ca Select Administrator: Command Prompt

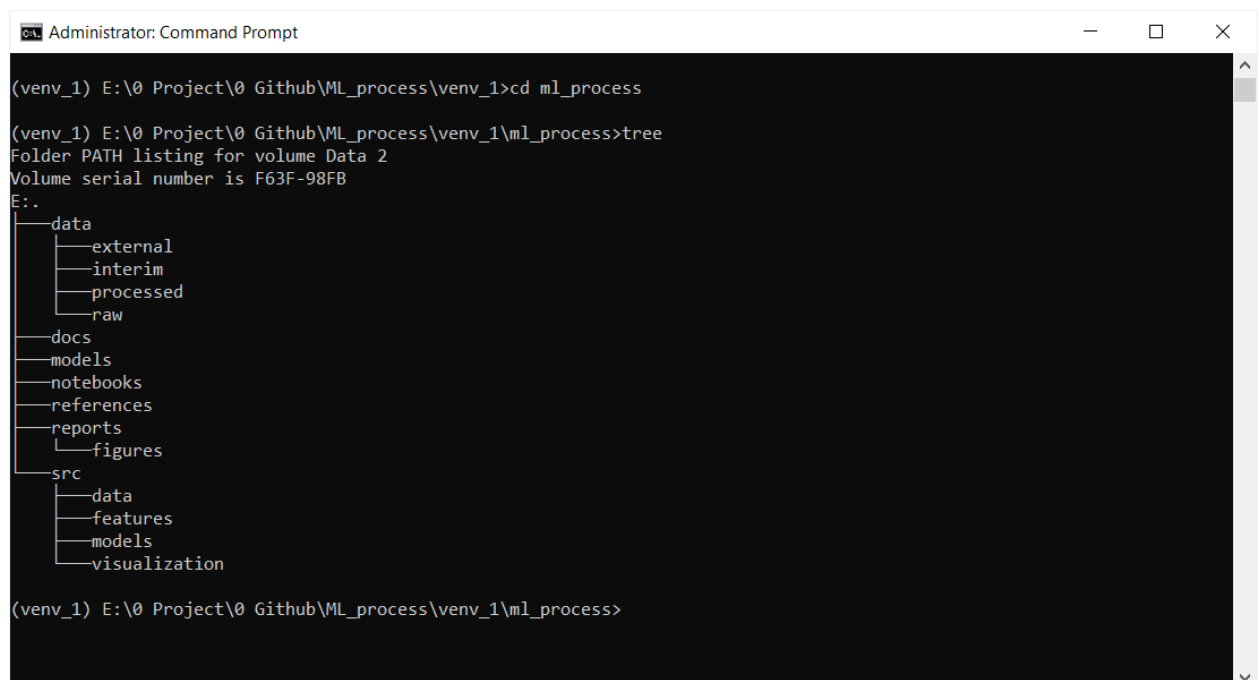
E:\0 Project\0 Github\ML_process\venv_1>.scripts\activate

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>dir
Volume in drive E is Data 2
Volume Serial Number is F63F-98FB

Directory of E:\0 Project\0 Github\ML_process\venv_1

28/02/2023  01:20    <DIR>          .
28/02/2023  01:20    <DIR>          ..
27/02/2023  23:24    <DIR>          Include
27/02/2023  23:24    <DIR>          Lib
28/02/2023  01:20    <DIR>          ml_process
27/02/2023  23:24                167 pyenv.cfg
28/02/2023  00:36    <DIR>          Scripts
                   1 File(s)                167 bytes
                   6 Dir(s)  54.199.492.608 bytes free

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>
```



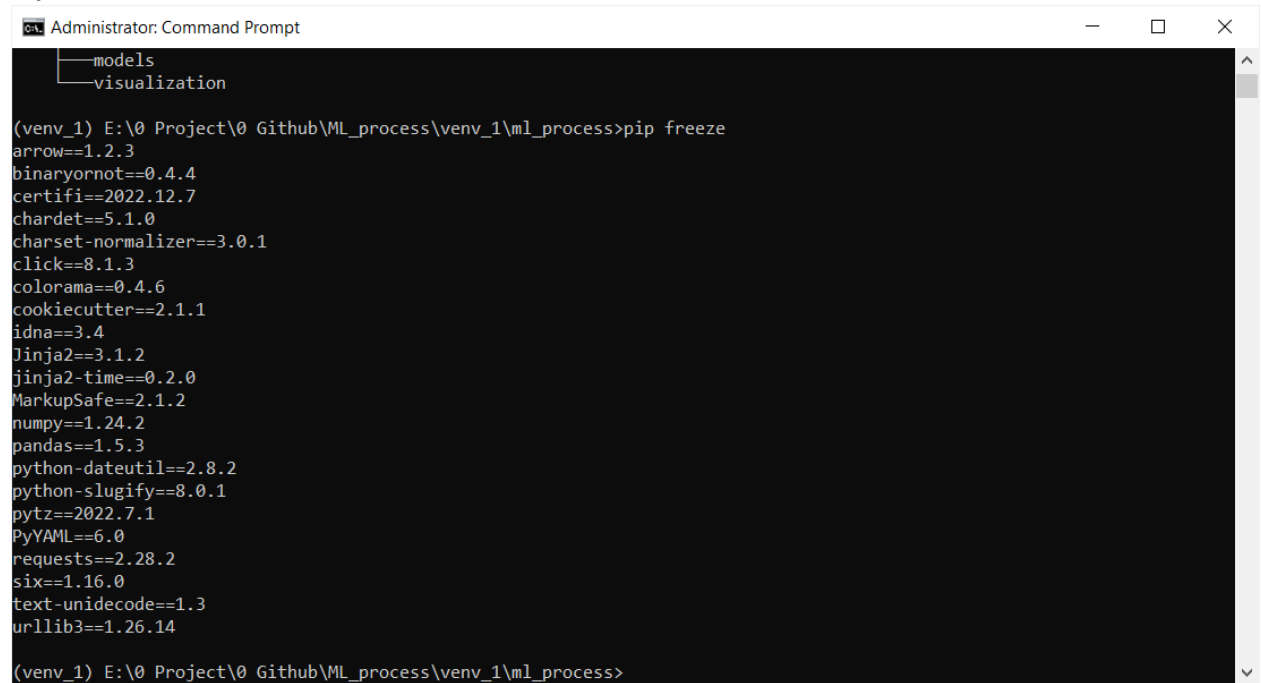
```
ca Administrator: Command Prompt

(venv_1) E:\0 Project\0 Github\ML_process\venv_1>cd ml_process

(venv_1) E:\0 Project\0 Github\ML_process\venv_1\ml_process>tree
Folder PATH listing for volume Data 2
Volume serial number is F63F-98FB
E:.
|-- data
|   |-- external
|   |-- interim
|   |-- processed
|   |-- raw
|-- docs
|-- models
|-- notebooks
|-- references
|-- reports
|   |-- figures
|-- src
|   |-- data
|   |-- features
|   |-- models
|   |-- visualization

(venv_1) E:\0 Project\0 Github\ML_process\venv_1\ml_process>
```

- **Step 4:**

**Pip Freeze**

```
Administrator: Command Prompt

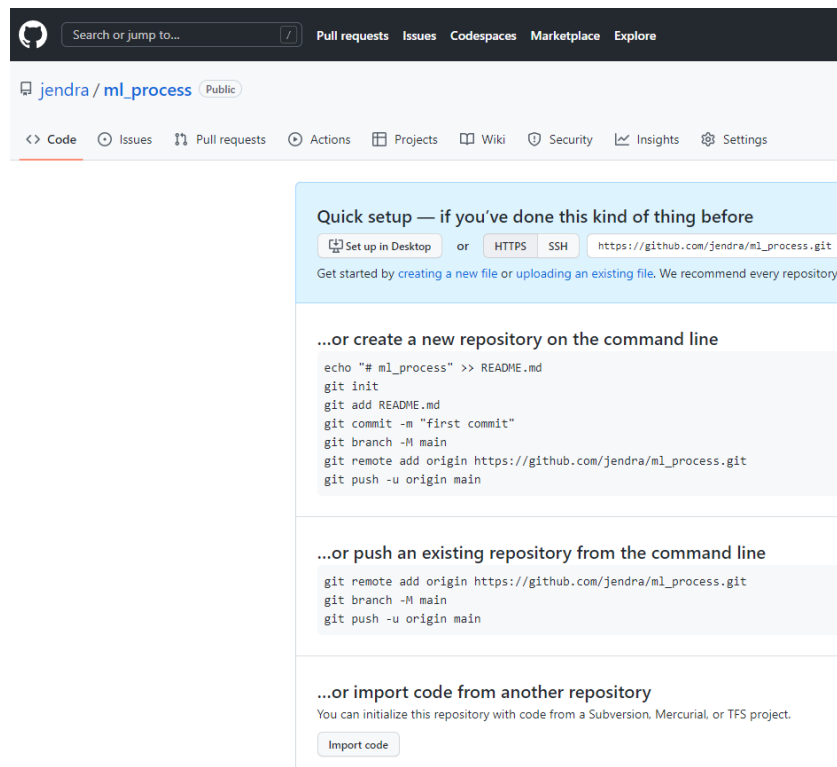
models
visualization

(venv_1) E:\0 Project\0 Github\ML_process\venv_1\ml_process>pip freeze
arrow==1.2.3
binaryornot==0.4.4
certifi==2022.12.7
chardet==5.1.0
charset-normalizer==3.0.1
click==8.1.3
colorama==0.4.6
cookiecutter==2.1.1
idna==3.4
Jinja2==3.1.2
jinja2-time==0.2.0
MarkupSafe==2.1.2
numpy==1.24.2
pandas==1.5.3
python-dateutil==2.8.2
python-slugify==8.0.1
pytz==2022.7.1
PyYAML==6.0
requests==2.28.2
six==1.16.0
text-unidecode==1.3
urllib3==1.26.14

(venv_1) E:\0 Project\0 Github\ML_process\venv_1\ml_process>
```

**3. Create repository account**

[https://github.com/jendra/ml\\_process](https://github.com/jendra/ml_process)



Quick setup — if you've done this kind of thing before

[Set up in Desktop](#) or [HTTPS](#) [SSH](#) [https://github.com/jendra/ml\\_process.git](https://github.com/jendra/ml_process.git)

Get started by [creating a new file](#) or [uploading an existing file](#). We recommend every repository

...or create a new repository on the command line

```
echo "# ml_process" >> README.md
git init
git add README.md
git commit -m "first commit"
git branch -M main
git remote add origin https://github.com/jendra/ml_process.git
git push -u origin main
```

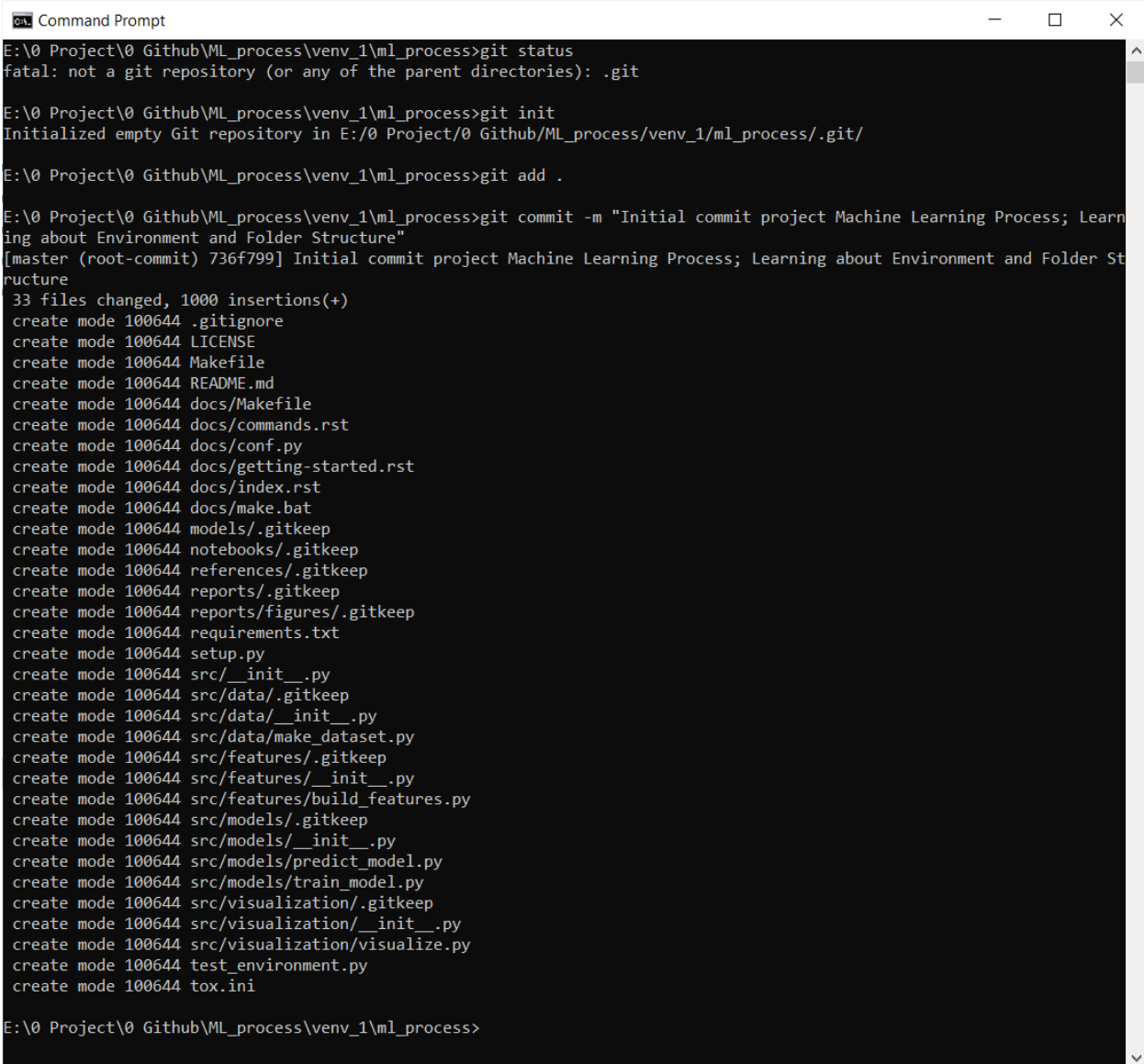
...or push an existing repository from the command line

```
git remote add origin https://github.com/jendra/ml_process.git
git branch -M main
git push -u origin main
```

...or import code from another repository

You can initialize this repository with code from a Subversion, Mercurial, or TFS project.

[Import code](#)

**4. Create local repository, add readme.md with body of your name and what project name.**

```
Command Prompt
E:\0 Project\0 Github\ML_process\venv_1\ml_process>git status
fatal: not a git repository (or any of the parent directories): .git

E:\0 Project\0 Github\ML_process\venv_1\ml_process>git init
Initialized empty Git repository in E:\0 Project\0 Github\ML_process\venv_1\ml_process/.git/

E:\0 Project\0 Github\ML_process\venv_1\ml_process>git add .

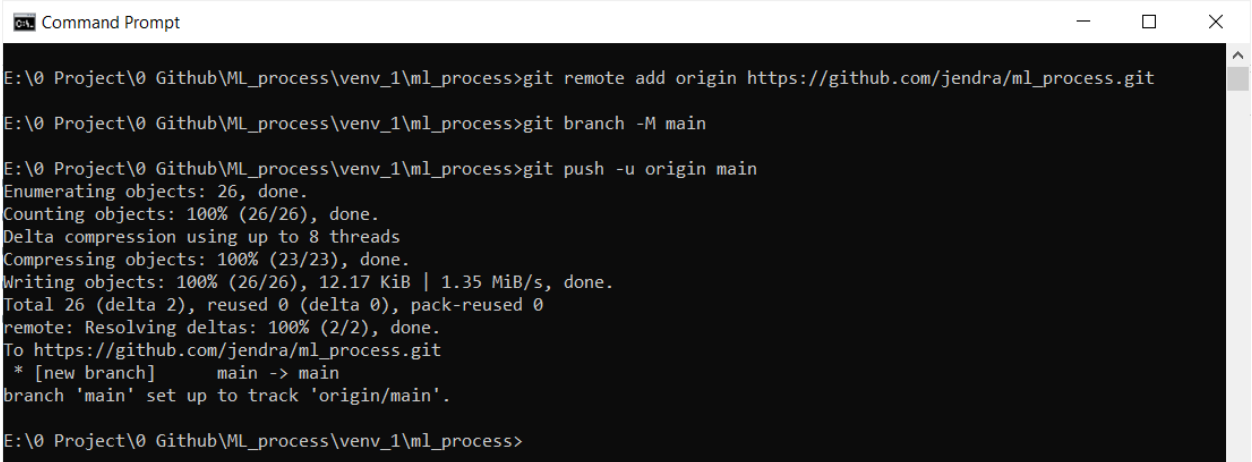
E:\0 Project\0 Github\ML_process\venv_1\ml_process>git commit -m "Initial commit project Machine Learning Process; Learning about Environment and Folder Structure"
[master (root-commit) 736f799] Initial commit project Machine Learning Process; Learning about Environment and Folder Structure
33 files changed, 1000 insertions(+)
create mode 100644 .gitignore
create mode 100644 LICENSE
create mode 100644 Makefile
create mode 100644 README.md
create mode 100644 docs/Makefile
create mode 100644 docs/commands.rst
create mode 100644 docs/conf.py
create mode 100644 docs/getting-started.rst
create mode 100644 docs/index.rst
create mode 100644 docs/make.bat
create mode 100644 models/.gitkeep
create mode 100644 notebooks/.gitkeep
create mode 100644 references/.gitkeep
create mode 100644 reports/.gitkeep
create mode 100644 reports/figures/.gitkeep
create mode 100644 requirements.txt
create mode 100644 setup.py
create mode 100644 src/__init__.py
create mode 100644 src/data/.gitkeep
create mode 100644 src/data/__init__.py
create mode 100644 src/data/make_dataset.py
create mode 100644 src/features/.gitkeep
create mode 100644 src/features/__init__.py
create mode 100644 src/features/build_features.py
create mode 100644 src/models/.gitkeep
create mode 100644 src/models/__init__.py
create mode 100644 src/models/predict_model.py
create mode 100644 src/models/train_model.py
create mode 100644 src/visualization/.gitkeep
create mode 100644 src/visualization/__init__.py
create mode 100644 src/visualization/visualize.py
create mode 100644 test_environment.py
create mode 100644 tox.ini

E:\0 Project\0 Github\ML_process\venv_1\ml_process>
```

**5. Create file with extension pdf that proofing you're already doing point 1-4.**

This pdf file will be uploaded in:

[https://github.com/jendra/ml\\_process](https://github.com/jendra/ml_process)

**6. Commit and push it to online repository.**

```
Command Prompt

E:\0 Project\0 Github\ML_process\venv_1\ml_process>git remote add origin https://github.com/jendra/ml_process.git

E:\0 Project\0 Github\ML_process\venv_1\ml_process>git branch -M main

E:\0 Project\0 Github\ML_process\venv_1\ml_process>git push -u origin main
Enumerating objects: 26, done.
Counting objects: 100% (26/26), done.
Delta compression using up to 8 threads
Compressing objects: 100% (23/23), done.
Writing objects: 100% (26/26), 12.17 KiB | 1.35 MiB/s, done.
Total 26 (delta 2), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (2/2), done.
To https://github.com/jendra/ml_process.git
 * [new branch]      main -> main
branch 'main' set up to track 'origin/main'.

E:\0 Project\0 Github\ML_process\venv_1\ml_process>
```

**Thank you very much. Any concerns/feedback please let me know.**