

# 실습 셋팅

교육 2 일 차

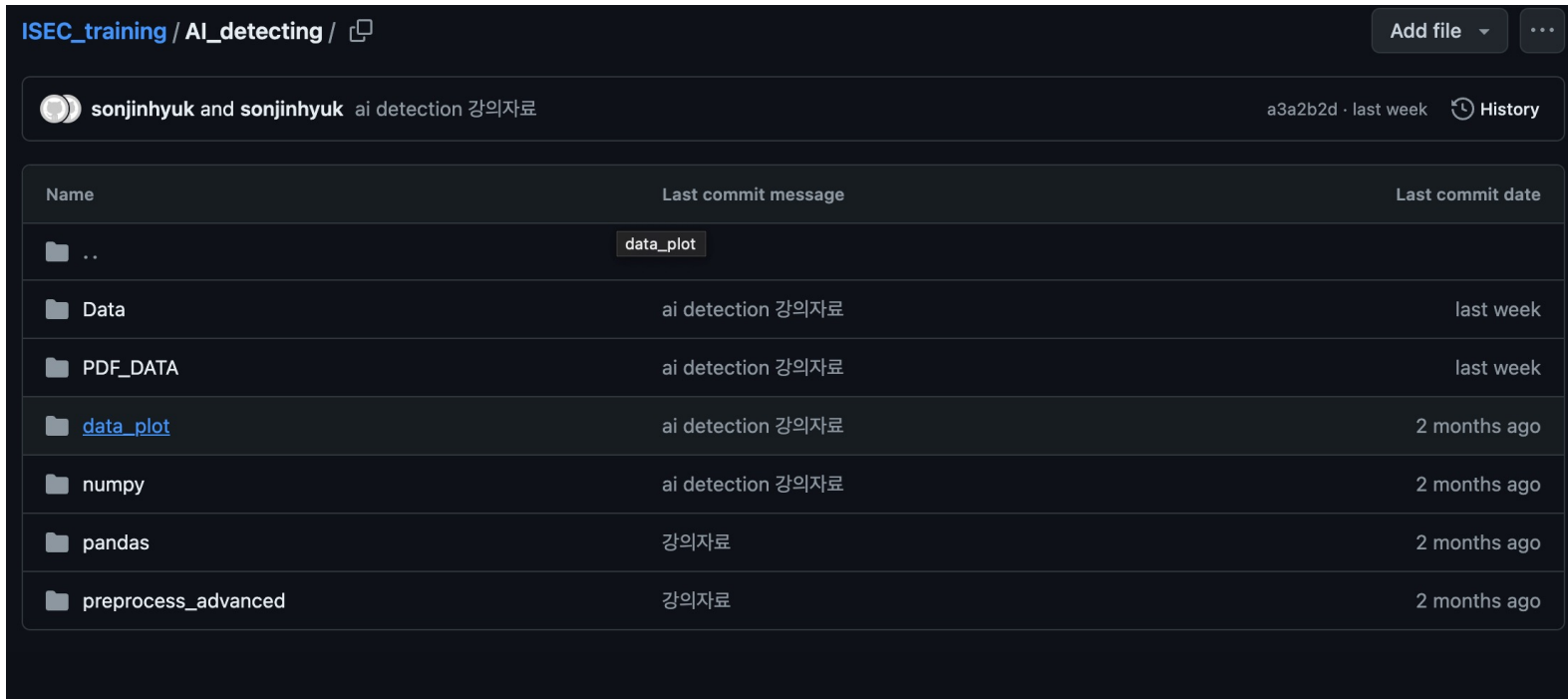
# Window(가상머신) 설정

○ 교육관련 자료 : Github ([https://github.com/sonjinhyuk/ISEC\\_training/AI\\_detecting](https://github.com/sonjinhyuk/ISEC_training/AI_detecting))

▪ 가상머신 : Oracle VirtualBox Image (USB 배포)

▪ Windows 10

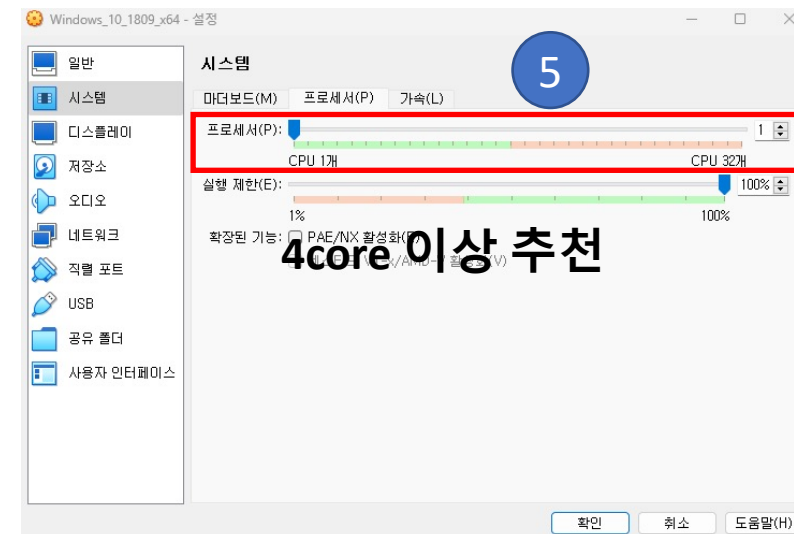
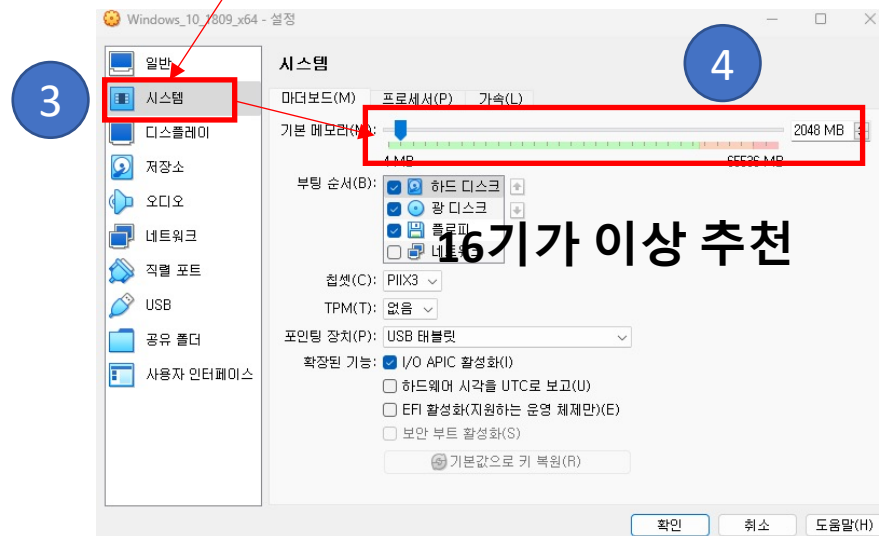
▪ Python 3.10

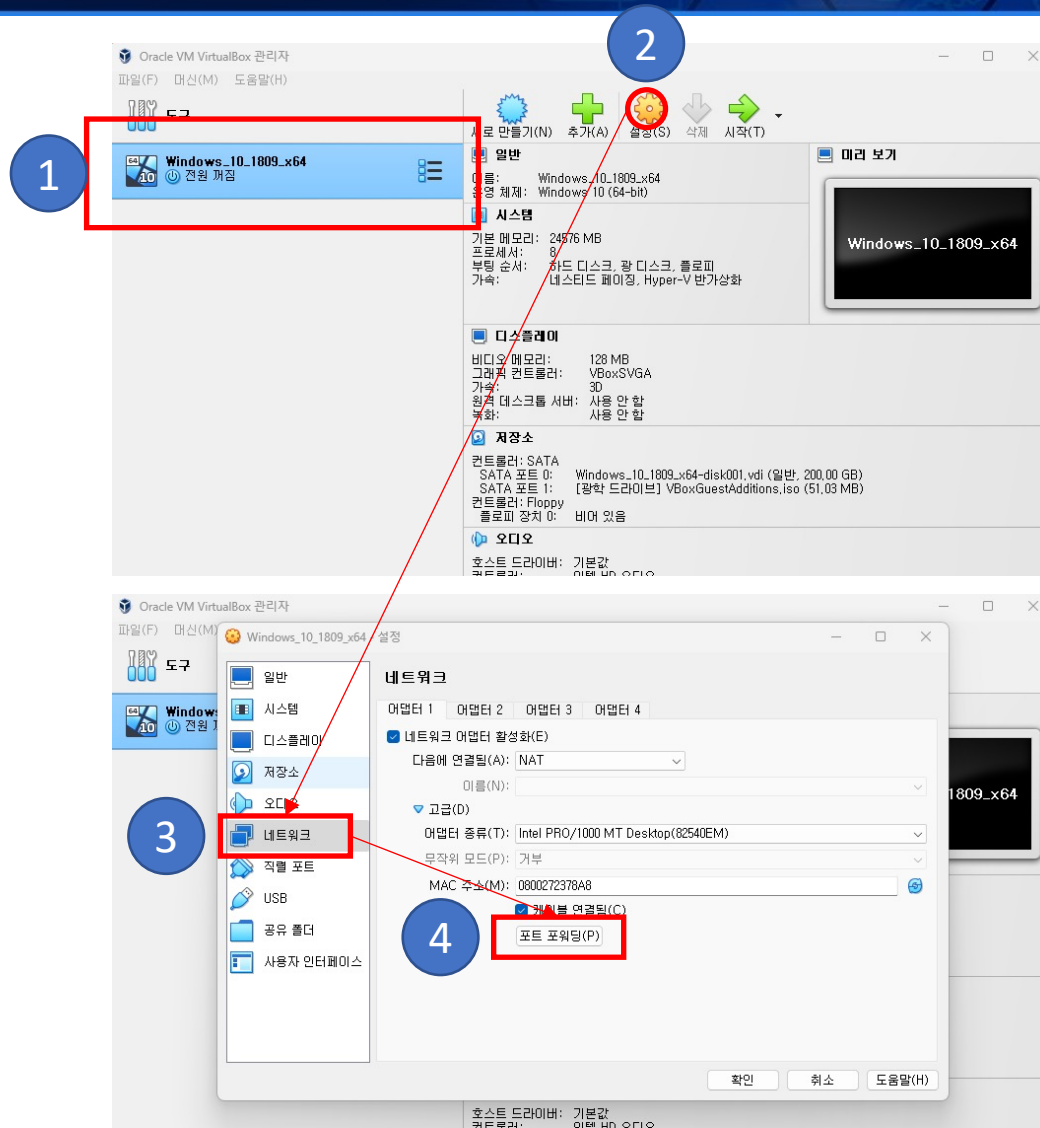


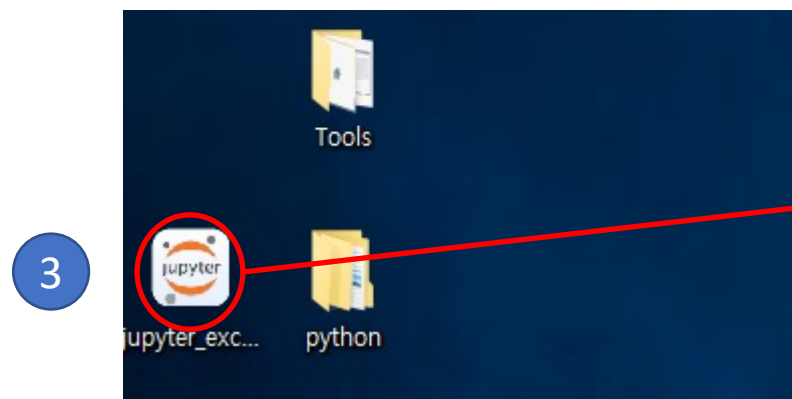
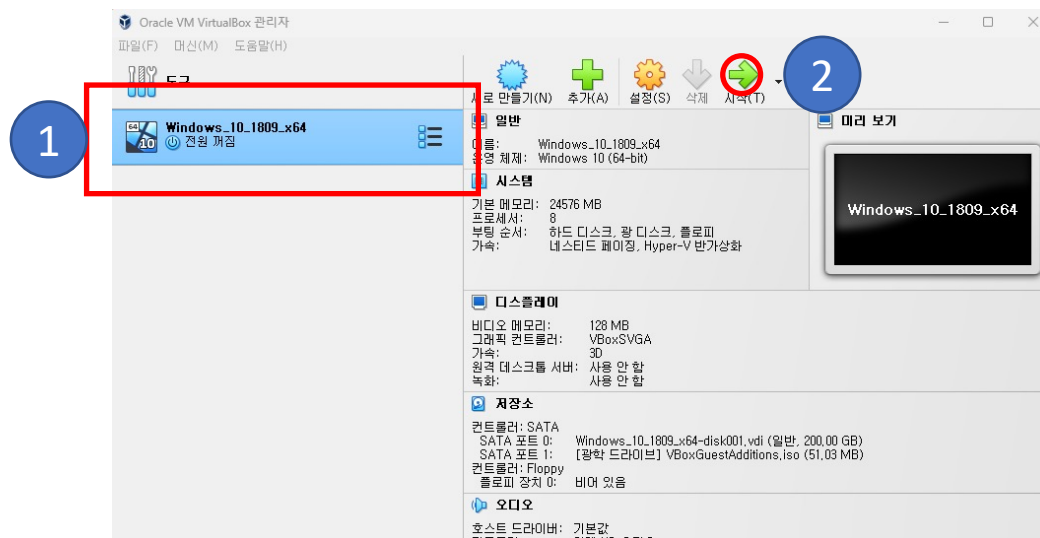
ISEC\_training / AI\_detecting

sonjinhyuk and sonjinhyuk ai detection 강의자료 a3a2b2d · last week History

Name	Last commit message	Last commit date
..	data_plot	
Data	ai detection 강의자료	last week
PDF_DATA	ai detection 강의자료	last week
<a href="#">data_plot</a>	ai detection 강의자료	2 months ago
numpy	ai detection 강의자료	2 months ago
pandas	강의자료	2 months ago
preprocess_advanced	강의자료	2 months ago

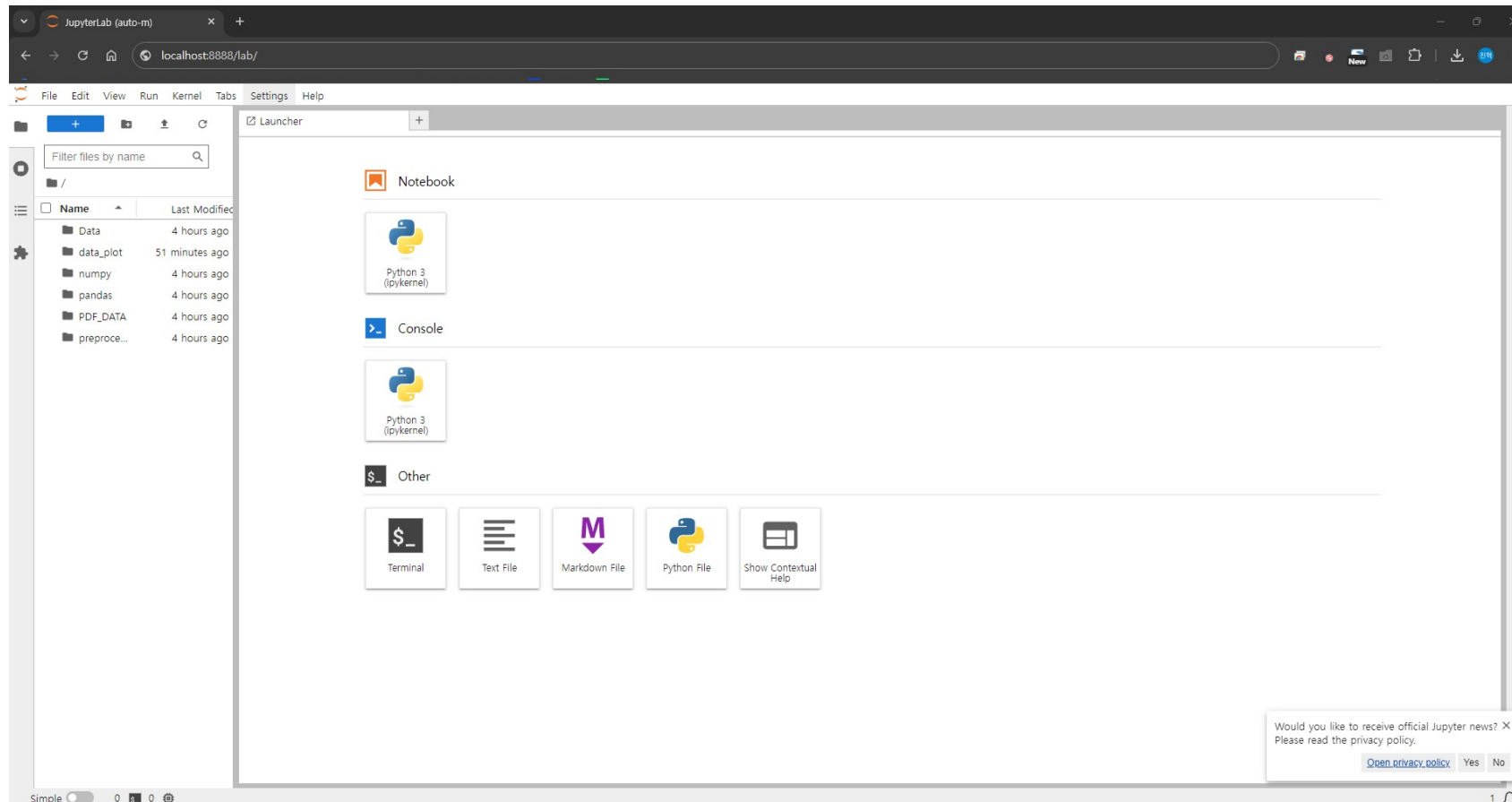






```
C:\Users\malware\Desktop\jupyter_excute.exe

[W 2024-07-19 15:19:05.613 ServerApp] notebook_dir is deprecated, use root_dir
[I 2024-07-19 15:19:05.613 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-07-19 15:19:05.620 ServerApp] notebook | extension was successfully linked.
[I 2024-07-19 15:19:05.663 ServerApp] notebook_shim | extension was successfully linked.
[W 2024-07-19 15:19:06.029 ServerApp] WARNING: The Jupyter server is listening on all IP addresses and not using encryption. This is not recommended.
[W 2024-07-19 15:19:06.029 ServerApp] WARNING: The Jupyter server is listening on all IP addresses and not using authentication. This is highly insecure and not recommended.
[I 2024-07-19 15:19:06.030 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-07-19 15:19:06.032 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-07-19 15:19:06.032 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-07-19 15:19:06.034 LabApp] JupyterLab extension loaded from C:\Users\malware\Desktop\python\AI\venv\lib\site-packages\jupyterlab
[I 2024-07-19 15:19:06.034 LabApp] JupyterLab application directory is C:\Users\malware\Desktop\python\AI\venv\share\jupyterlab
[I 2024-07-19 15:19:06.035 LabApp] Extension Manager is 'pip'.
[I 2024-07-19 15:19:06.046 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-07-19 15:19:06.049 ServerApp] notebook | extension was successfully loaded.
[I 2024-07-19 15:19:06.049 ServerApp] Serving notebooks from local directory: C:\Users\malware\Desktop\python\AI\AI_detection
[I 2024-07-19 15:19:06.049 ServerApp] Jupyter Server 2.14.0 is running at:
[I 2024-07-19 15:19:06.049 ServerApp] http://localhost:8888/lab
[I 2024-07-19 15:19:06.050 ServerApp] http://127.0.0.1:8888/lab
[I 2024-07-19 15:19:06.050 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[I 2024-07-19 15:19:06.089 ServerApp] Skipped non-installed server(s): bash-language-server, dockerfile-language-server-nodejs, javascript-typescript-langserver, jedi-language-server, julia-language-server, pyright, python-language-server, python-lsp-server, r-languageserver, sql-language-server, texlab, typescript-language-server, unified-language-server, vscode-css-languageserver-bin, vscode-html-languageserver-bin, vscode-json-languageserver-bin, yamll-language-server
```



○ 노트북 브라우저 실행

○ <https://localhost:8888/lab> 접속





The screenshot displays a Jupyter Notebook environment. On the left, a file browser shows the directory structure: / Data / test\_files (2 days ago) and pdf\_extraction.ipynb (13 seconds ago, 3.1 KB). The file pdf\_extraction.ipynb is selected. The main area shows the notebook content, which includes a code cell with the following text:

```
#PDF 파일을 활용하여 csv 파일을 생성
```

- parser import

```
[1]: import sys
sys.path.append("C:\\Users\\malware\\Desktop\\python\\AI\\Docscanner_parser\\docscanner_parser")
from CSRC_parser.PDFparser import preprocessing
```

- file load

```
[2]: from pathlib import Path
from tqdm import tqdm
paths = Path("Data\\test_files").glob("**")
results = []
for p in tqdm(paths):
    try:
        results.append(preprocessing.pdfparser(str(p)))
    except:
        print(p)

@it [00:00, ?it/s]
```

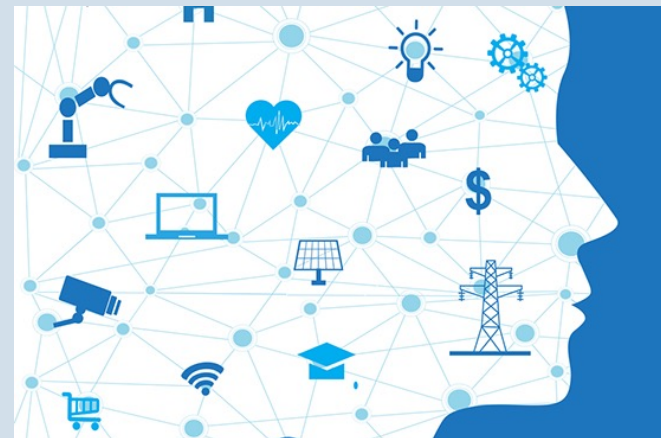
```
[3]: for series in results:
    file_name = series.model_input['file_name'].split("\\")[-1]
    series.model_input.to_csv(f"Data\\{file_name}.csv")
```

The code cell is highlighted with a red box, and the output area shows a progress bar and a message: @it [00:00, ?it/s].



# MAC 사용자 설정

1. python setting
  1. `python -m venv venv`
  2. `pip install -r requirements.txt`
2. Git install
  1. `git --version`
3. Nodejs install
  1. `brew install node`



# 감사합니다

CYBER SECURITY RESEARCH CENTER