

# Machine Learning Engineering Program

*Week 2 Data Collection & Extraction*

Jannie Li

# Use Tesseract with Python

Input: Image.png



Output:

```
output.txt
File Edit View
El café El batido La cerveza La leche
El champán El agua El vino El zumo
```

Python:

```
tess.py
File Edit View
from PIL import Image
import pytesseract
import os

image_path = 'processed.png'

if not os.path.exists(image_path):
    print(f"Image not found: {image_path}")
else:
    image = Image.open(image_path)
    text = pytesseract.image_to_string(image)
    if text.strip():
        print("OCR output:")
        print(text)
    else:
        print("No text detected in the image.")
```

```
(venv) jannie@DESKTOP-DAIMEU6:/mnt/c/Users/janni/Desktop/Inference/assignment 2$ tesseract image.png output -l spa
Estimating resolution as 408
Detected 12 diacritics
```

# Thank you!

[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Jannie Li Discord: fancyavatar777\_10384