

Panduan Instalasi Requirement

- Tesseract OCR
- Poppler
- Git Bash

Langkah-langkah Instalasi:

1. Instalasi Tesseract OCR

- Unduh Tesseract OCR dari situs resmi
- Jalankan installer dan ikuti petunjuk instalasi
- Secara default, Tesseract akan terinstal di C:\Program Files\Tesseract-OCR

2. Konfigurasi Poppler

- Unduh folder Poppler
- Salin folder Poppler ke drive C atau D pada komputer Anda
- Contoh lokasi: C:\poppler atau D:\poppler

3. Konfigurasi Environment Variable Untuk Tesseract:

- Buka "Environment Variables" (klik kanan pada This PC/My Computer → Properties → Advanced System Settings → Environment Variables)
- Pada bagian "User variables", cari variabel "Path"
- Klik "Edit"
- Klik "New" dan tambahkan path instalasi Tesseract (contoh: C:\Program Files\Tesseract-OCR)
- Klik "OK"

Untuk Poppler:

- Pada bagian "System variables", cari variabel "Path"
- Klik "Edit"
- Klik "New" dan tambahkan path folder bin Poppler (contoh: C:\poppler\Library\bin)
- Klik "OK" untuk menyimpan perubahan

4. Verifikasi Instalasi

- Buka Command Prompt baru
- Ketik `tesseract --version` untuk memverifikasi instalasi Tesseract
- Sekarang Python script Anda siap dijalankan dengan library yang membutuhkan Tesseract dan Poppler

5. Verifikasi Database

- Akses directory `database_pegawai`
- Rename file `template_database_pegawai_main.xlsx` menjadi `database_pegawai_main.xlsx`
- isi data di file tersebut lalu save

6. Pastikan file tersedia di directory `input`

7. Pastikan tidak terdapat directory `cropped_images`

Fix Common Issue:

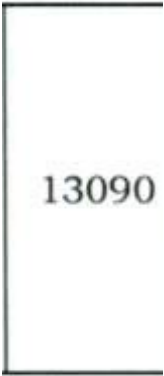
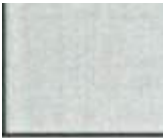
Hasil OCR tidak sesuai:

Untuk keperluan debugging bisa comment `#` di bagian `# Menyelesaikan file sementara dan menghapusnya` contoh:

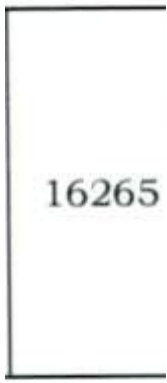
```
# Menyelesaikan file sementara dan menghapusnya
# print(
#     f"PDF dan gambar berhasil dibuat untuk file: {text_nip}__{satker_pegawai}_"
#     os.remove(output_image_path_nip)
# # write log NIP tidak ditemukan
# if nip_not_found:
#     with open(os.path.join(current_directory, "output", file_name, "nip_not_found.
#         for nip in nip_not_found:
#             log_file.write(f"NIP tidak ditemukan: {nip}\n")
# os.rmdir(output_crop_dir)
```

Pastikan image yang di crop tidak terdapat 2 atau lebih garis horizontal diatas angka seperti ini:

Gambar yang salah



Gambar yang benar



Cara Penanganan

1. Akses file `pdf_processor` pergi ke bagian # Tentukan koordinat crop berdasarkan jenis pemindahan sesuaikan koordinat yang ada