PROJECT DOCUMENTATION

Job Application & CV Parser

with OCR Automation & n8n Integration

Generated Date:	January 31, 2026
Version:	1.0.0
Author:	Dilawar Shah
Technology:	Python, FastAPI, Streamlit, n8n

1. Executive Overview

This project is a full-stack automated application tracking system (ATS) starter. It features a user-friendly frontend for submitting job applications and a robust backend that performs Optical Character Recognition (OCR) on resumes (PDF, Word, or Images) before sending the data to an n8n workflow for downstream processing.

1.1 Key Features

- **Universal File Support:** Accepts PDF, DOCX, PNG, JPG.
- **Intelligent Pre-processing:** Converts files to high-contrast grayscale for OCR.
- **Rapid OCR:** Uses RapidOCR (ONNX) for fast text extraction.
- **Automation Ready:** Instant integration with n8n Webhooks.

2. System Architecture

Component	Technology	Description
Frontend	Streamlit	User interface for file uploads.
Backend API	FastAPI	Handles routing and file processing.
OCR Engine	RapidOCR	Extracts text from images.
Image Proc	OpenCV / Pillow	Grayscale conversion & resizing.
Automation	n8n	Workflow orchestration (Email, Sheets).

3. Installation & Setup

Step 1: Create Environment

```
conda create --name ocr-env python=3.11 -y
conda activate ocr-env
```

Step 2: Install Dependencies

Ensure you have the following requirements:

```
pip install streamlit fastapi uvicorn requests pymupdf
pip install rapidocr_onnxruntime opencv-python pillow docx2pdf
```

4. Running the Application

The system requires two separate terminals:

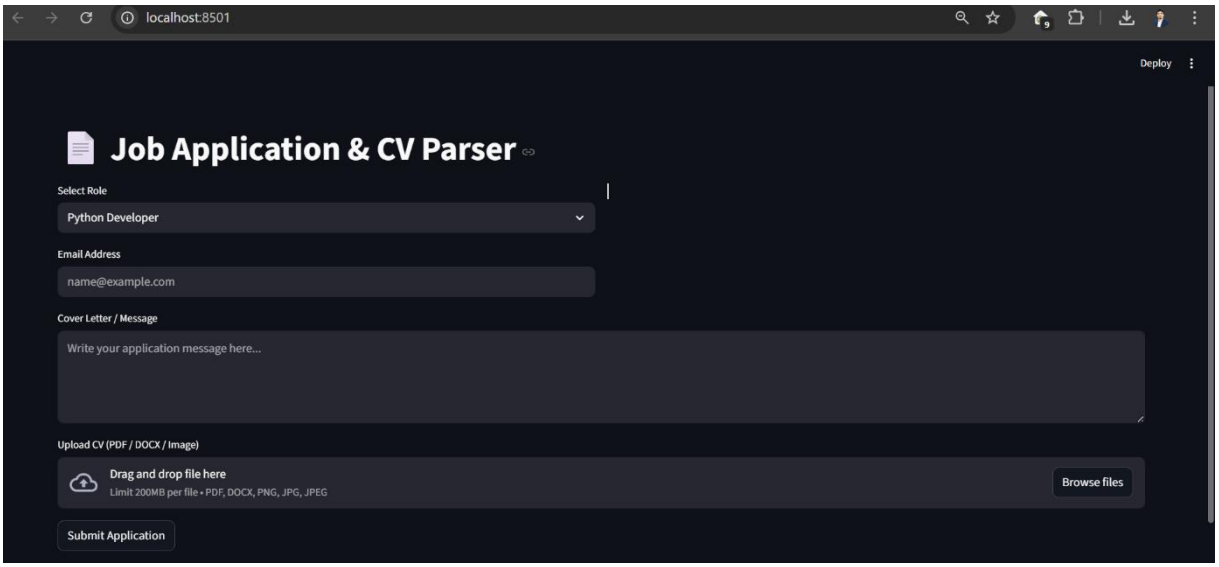
Terminal A: Backend API

```
uvicorn main:app --reload
```

Terminal B: Frontend UI

```
streamlit run app.py
```

Frontend UI



N8N Workflow