CV_Handler Documentation

1. Project Overview

CV_Handler is a cloud-based document management and analytics system. It allows users to upload, search, sort, and classify large volumes of PDF and DOCX files. It uses Supabase as the cloud backend and is developed in Python.

2. Features

- Upload PDF and DOCX documents to the cloud
- Extract and sort documents by actual title text
- Search within documents for keywords and highlight matches
- Classify documents based on predefined categories
- View analytics: number of documents, storage size, processing time

3. Tech Stack

- Python (core logic and text processing)
- Supabase (PostgreSQL + file storage)
- pdfminer.six / python-docx for document parsing
- Scikit-learn (optional) for classification logic
- Jupyter Notebook for prototyping and demonstration

4. Usage

1. Clone the GitHub repository
2. Open the Jupyter notebook and run the cells
3. Upload documents or connect to a Supabase storage bucket
4. Use the provided functions to search, sort, and classify
5. View or export results and statistics

5. Algorithm Design

Sorting is done by extracting the title text from documents using PDF and DOCX parsers. Search is implemented through keyword matching and optional regex. Classification uses either keyword-rule based or machine learning-based (e.g. Naive Bayes) approaches, depending on user configuration.

6. Cloud Setup (Supabase)

The system connects to Supabase via API keys and URLs stored in environment variables. Files are uploaded to Supabase storage, and metadata is optionally stored in the database. Authentication and permissions can be managed through Supabase dashboard.

7. How to Run

You can run this project locally by opening the Jupyter Notebook. Make sure to install the required libraries listed in `requirements.txt`. If hosting online, you can adapt the code into a Streamlit or Gradio app.