SMART RESUME PARSER

# Project Title

Smart Resume Parser

# Introduction

Hiring managers often deal with hundreds of resumes in various formats. Manually extracting key information like email, phone number, and skills is time-consuming. This project aims to automate that process using Natural Language Processing (NLP).

# Abstract

The Smart Resume Parser is a Streamlit-based Python application that extracts structured information from resumes in PDF and DOCX formats. It uses spaCy for Named Entity Recognition (NER), along with regular expressions, to identify important data like name, email, phone number, and skills. Users can upload a resume and receive a downloadable CSV file containing the extracted data. This tool helps recruiters and HR professionals streamline the resume screening process.

# Tools & Technologies Used

- Python 3.11+  
- spaCy  
- PyMuPDF (fitz)  
- python-docx  
- Streamlit  
- pandas  
- regex (re)

# Steps Involved in Building the Project

1. Project Setup:  
 - Created virtual environment and installed dependencies (streamlit, spacy, pymupdf, python-docx, pandas)  
 - Downloaded en\_core\_web\_sm spaCy model  
  
2. Text Extraction Logic:  
 - Used PyMuPDF for extracting text from PDFs  
 - Used python-docx for extracting text from Word files  
  
3. Information Parsing:  
 - Used spaCy to extract person names  
 - Used regex to extract emails and phone numbers  
 - Looked for "Skills" section using keyword matching  
  
4. Streamlit UI Development:  
 - Created a file uploader  
 - Displayed extracted data in JSON format  
 - Enabled CSV download of parsed results  
  
5. Testing and Sample Files:  
 - Added 3 sample resumes (PDF & DOCX) for testing  
 - Stored test results in data/outputs/ folder  
  
6. Exporting & Finalization:  
 - Generated clean CSV output  
 - Saved results to local folder  
 - Prepared the app for GitHub and cloud deployment

# Conclusion

This project allowed me to gain hands-on experience with real-world file processing, NLP, and frontend-backend integration using Python. It improved my understanding of text extraction, pattern recognition, and data visualization with Streamlit. The Smart Resume Parser is a practical HR tech tool and a solid addition to my full stack development portfolio.