Resume Parser Using NLP & Streamlit

A Python-Based NLP Project

Presented By
Mayank Sharma

Aspiring Data Scientist | NLP Enthusiast

📌 Project Overview

The Resume Parser using NLP & Streamlit is a Python-based application designed to automatically extract key information from resumes in PDF format. The project demonstrates the practical use of Natural Language Processing (NLP) techniques to process unstructured text data and convert it into meaningful structured information.

This application provides an interactive Streamlit web interface, allowing users to upload resumes and instantly view extracted details.

🎯 Objective of the Project

The main objectives of this project are:

- To understand real-world applications of Natural Language Processing
- To extract structured data from unstructured resume text
- To build an end-to-end NLP-based web application
- To showcase practical skills suitable for industry and academic purposes

⚙️ Key Features

Upload resumes in PDF format

Extract important details such as:

- Name
- Email address
- Phone number
- Skills

Uses NLP and Regular Expressions for text extraction

Extract text from the Resume(pdf)

[2]:
```python
import pdfplumber

def extract_text(path):
    text = ""
    with pdfplumber.open(path) as pdf:
        for page in pdf.pages:
            text += page.extract_text()
    return text
```

Clean Extracted Text

[3]:
```python
import re

def clean_text(text):
    # Remove bullet symbols like \x7f
    text = re.sub(r'\x7f', ' ', text)

    # Replace newlines with space
    text = re.sub(r'\n', ' ', text)

    # Remove standalone numbers (list artifacts)
    text = re.sub(r'\b\d+\b', ' ', text)

    # Normalize all whitespace
    text = re.sub(r'\s+', ' ', text)

    return text.strip()
```

Extract Email & Number

[4]:
```
def extract_email(text):
    emails= re.findall(r'\S+@+\S+', text)
    return emails[0] if emails else None


def extract_number(text):
    phone_pattern = r'(\+?\d{1,3}[-\s]?)?\d{5}[-\s]?\d{5}'
    phones = re.findall(phone_pattern, text)
    return phones[0] if phones else "Not found"
```

Extract Name

[5]:
```
import spacy
nlp=spacy.load('en_core_web_sm')
```

[6]:
```
def extract_name(text):
    doc=nlp(text)
    for ent in doc.ents:
        if ent.label_=='PERSON':
            return ent.text
    return None
```

Extract Skills

[7]:
```
def extract_skills(text):
    with open('skills.txt') as f:
        skills=[s.strip().lower() for s in f.readlines()]

    text=text.lower()
    return [skill for skill in skills if skill in text]
```

Final Pipeline

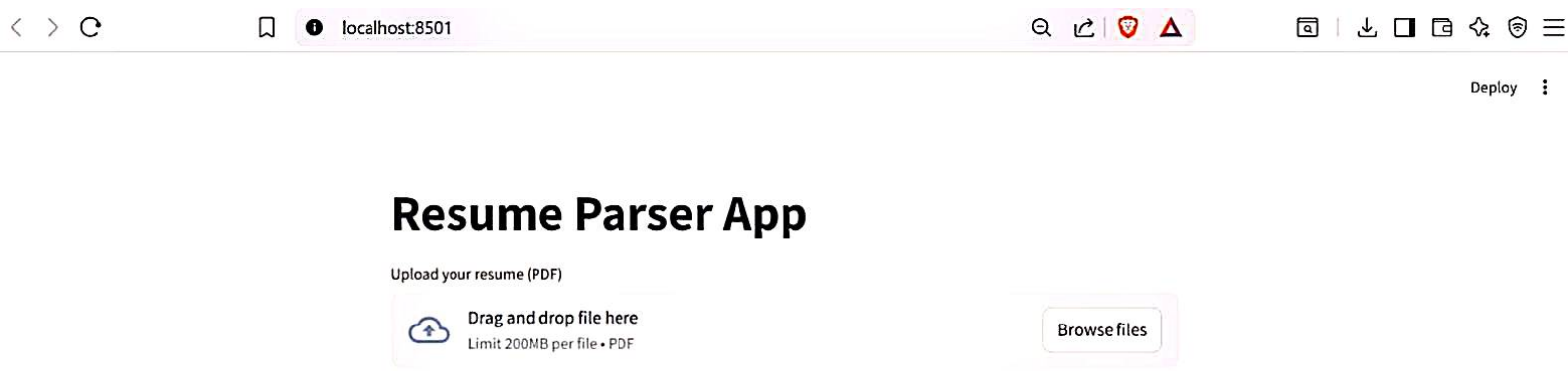
[8]:
```
def resume_parser(path):
    text = extract_text(path)
    text = clean_text(text)

    return {
        "Name": extract_name(text),
        "Email": extract_email(text),
        "Phone": extract_number(text),
        "Skills": extract_skills(text),
    }
```

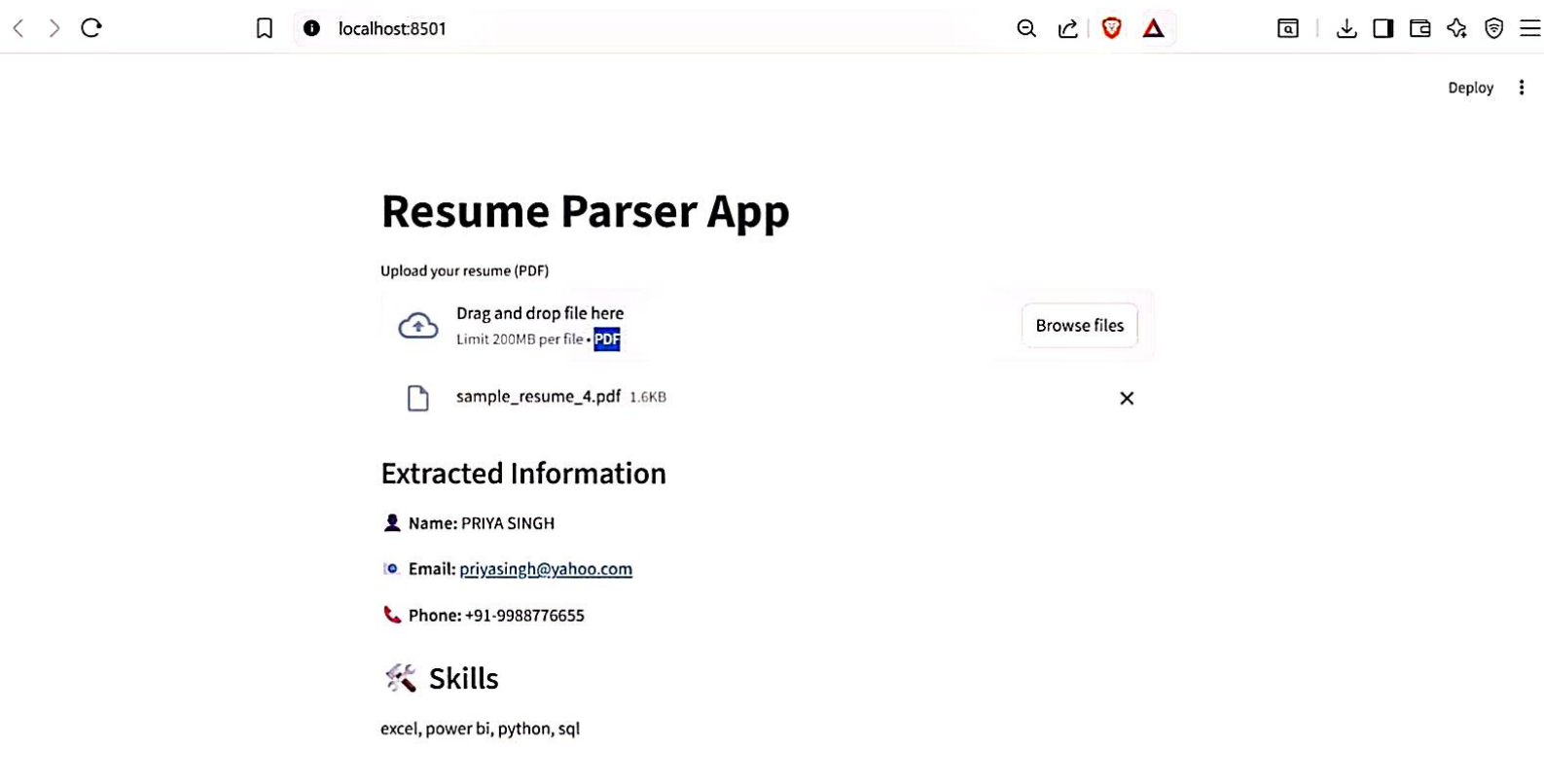
Test the Pipeline

[9]:
```
resume_parser('sample_resume_ml_training.pdf')
```

[9]:
```
{'Name': 'John Doe Data',
 'Email': 'johndoe@email.com',
 'Phone': 'Not found',
 'Skills': ['python', 'machine learning', 'sql', 'nlp']}
```

[ ]:


localhost:8501
Deploy
Resume Parser App
Upload your resume (PDF)
Drag and drop file here
Limit 200MB per file • PDF
Browse files


localhost:8501
Deploy
Resume Parser App
Upload your resume (PDF)
Drag and drop file here
Limit 200MB per file • PDF
Browse files
sample_resume_4.pdf 1.6KB
Extracted Information
Name: PRIYA SINGH
Email: priyasingh@yahoo.com
Phone: +91-9988776655
Skills
excel, power bi, python, sql