Machine Learning Project Report

Title of the Project:

AI Resume Parser Pro

Objective / Goal:

The objective of this project is to build an intelligent, web-based application capable of parsing resumes and extracting relevant professional and educational information using Natural Language Processing (NLP) techniques. The system aims to assist recruiters and HR professionals in streamlining candidate screening by automatically identifying key attributes such as personal details, technical skills, education background, work experience, and years of experience from uploaded PDF resumes.

Dataset Used:

This project does not rely on a static dataset. Instead, it processes resumes uploaded by users in real-time. The resumes are assumed to be in PDF format and follow a variety of formatting styles to simulate real-world usage. The application is designed to handle diverse resume structures to improve robustness and generalization.

Tools and Technologies Used:

- **Python**: Core programming language used for data processing and backend logic.
- **NLTK (Natural Language Toolkit)**: Employed for sentence tokenization and basic text processing.
- **PyMuPDF** (fitz): Used for extracting textual content from PDF files.
- **Streamlit**: Utilized to develop an interactive and responsive web-based user interface.
- **Regular Expressions (re)**: Applied to identify and extract structured data patterns such as email addresses, phone numbers, and names.
- **Pandas**: Used for structuring and exporting parsed information.
- **HTML/CSS (Custom Styling in Streamlit)**: Enhanced UI presentation with dynamic visual elements and responsive layouts.

Steps Performed:

1. Interface Development: A custom-designed web interface was developed using Streamlit, incorporating responsive UI components and styled using embedded CSS for a modern and user-friendly

experience.

2. Resume Upload: Users can upload their resumes in PDF format. A drag-and-drop or click-to-select interface facilitates easy uploads.

3. Text Extraction: Uploaded PDFs are processed using PyMuPDF to extract all textual content, which is then used for NLP-based parsing.

4. Information Extraction:

- Name, Email, and Phone Number: Extracted using regular expressions and positional heuristics from the top section of the resume.
- Technical Skills: Identified using keyword matching against a predefined categorized skills database.
- Education: Degree, field, and institution are extracted using NLP patterns and entity matching.
- Experience Sentences: Filtered from the text using key role-related phrases.
- Years of Experience: Estimated from quantified expressions like "5+ years of experience".

5. Data Presentation: Parsed data is visually rendered in card-style layouts using HTML/CSS. Summary metrics are shown in a dashboard view, and raw extracted text is available under an expandable section.

6. Download Functionality: Parsed resume data can be downloaded as a structured JSON file.

Result and Accuracy:

The application performs reliably on resumes with moderate-to-good formatting. It achieves approximately 80–90% accuracy in extracting key fields such as email, phone number, and technical skills, with slightly variable accuracy for education and experience extraction due to formatting inconsistencies in user-submitted resumes. User feedback and real-world testing continue to guide refinements.

Conclusion:

The AI Resume Parser Pro effectively demonstrates an end-to-end solution for resume parsing using modern web frameworks and NLP techniques. It combines usability with functional robustness, making it a valuable tool for HR teams and job application systems. Key learnings include designing flexible regex patterns, optimizing PDF text extraction, and using Streamlit for polished web deployment.

Future Work:

- Integration of Named Entity Recognition (NER) models to improve context-aware data extraction.
- Expansion to support DOCX and TXT formats.
- Addition of multilingual support.

- Implementation of a database to store and manage parsed resumes for batch processing and analytics.
- Collaboration with HR platforms to embed the parser as a plug-and-play module.