

# Smart Resume Parser Report

## Introduction

The Smart Resume Parser project focuses on automatically extracting structured information from resumes in PDF and DOCX formats. The goal is to reduce manual effort in resume screening by converting unstructured resume content into meaningful, machine-readable data such as skills, education, and work experience.

## Abstract

This project presents an intelligent resume parsing system built using Python and Natural Language Processing techniques. By extracting text from resumes and applying text preprocessing, the system identifies and organizes key resume sections. The parsed information is displayed through a simple user interface and can be exported for further analysis or storage.

## Tools Used

- Python – Core programming language for logic and data processing.
- spaCy – Used for Natural Language Processing and entity extraction.
- PyMuPDF – Used to extract text from PDF resumes.
- python-docx – Used to extract text from DOCX resumes.
- Streamlit – Used to build the interactive web-based user interface.

## Steps Involved in Building the Project

1. Accept resume uploads in PDF or DOCX format through the Streamlit interface.
2. Extract raw text from resumes using PyMuPDF or python-docx.
3. Clean and preprocess the extracted text by removing noise and normalizing content.
4. Apply spaCy and regular expressions to identify sections such as skills, education, and experience.
5. Organize the extracted information into structured formats like JSON or tables.
6. Provide options to export the parsed data as CSV or JSON files.

## Conclusion

The Smart Resume Parser project demonstrates the practical application of Natural Language Processing in real-world recruitment scenarios. It simplifies resume screening by converting unstructured documents into structured data efficiently. This project highlights strong skills in Python development, text processing, and building user-friendly analytical applications.