AI Skill Gap Analyzer

Document Processing Pipeline for Resume & Job Description Analysis - Infosys Springboard Internship 6.0

by Ravichandra D

Under Guidance of Praveen Sir

Contents

Project Introduction & Objectives

AI Skill Gap Analyzer Vision

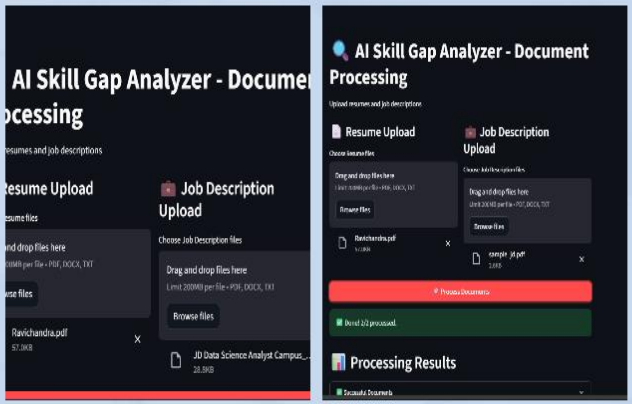
A Streamlit-based application designed to revolutionize recruitment by analyzing resumes against job descriptions to identify skill gaps and provide actionable recommendations.

Document Processing

Process multiple resume formats (PDF, DOCX, TXT) and job descriptions efficiently.

AI-Driven Analysis

Leverage machine learning to extract, clean, and structure text for skill gap identification.



System Architecture & Components

DocumentUploader

Handles file uploads via Streamlit, supports PDF/DOCX/TXT, validates size/type requirements.

TextExtractor

Extracts raw text using PyPDF2, python-docx, or encoding for TXT files with reliability.

DocumentProcessor

Main pipeline coordinating upload → extract → clean → display → download workflow.

TextCleaner

Cleans text artifacts, removes noise, structures sections, and ensures readability.

Complete Workflow Pipeline

Step 1: File Upload & Validation

User uploads resumes and job descriptions with automatic validation for file type and size limits.

Step 2: Text Extraction

Reliable text extraction using multiple methods to handle different document formats seamlessly.

Step 3: Text Cleaning & Structuring

Advanced cleaning removes artifacts, structures content into readable sections with proper formatting.

Step 4: Results Display

Real-time processing statistics and cleaned text preview within the Streamlit interface.

Step 5: Export & Download

Flexible export options in CSV and JSON formats for further analysis and integration.

Current Features & Capabilities

3+	2	5	100%
File Formats	**Export Options**	**Processing Steps**	**Real-time Preview**

Multi-file Upload

Simultaneous processing of resumes and job descriptions with intelligent validation.

- Support for PDF, DOCX, TXT formats
- Automatic file size validation
- Batch processing capabilities

AI Processing

Advanced text extraction and cleaning with structured section detection.

- Regex-based section detection
- Noise removal and formatting
- Progress tracking with status updates

Export & Integration

Flexible output formats for seamless integration with downstream systems.

- CSV format for spreadsheet analysis
- JSON for API integration
- Clean structured data output

Limitations & Future Improvements

Development Roadmap

Current limitations and strategic improvements planned for the AI Skill Gap Analyzer system.

Current Limitations

- Textract dependency installation issues
- Regex-based section detection limitations
- Large PDFs (>10MB) not supported
- Poor formatting affects cleaning accuracy

Future Enhancements

- OCR support for scanned PDFs
- ML-based section detection
- Resume-Job similarity analysis
- Advanced search and highlight features

Milestone 1 Complete

Data ingestion and parsing foundation established

Next Steps

Skill extraction, gap detection, and recommendation generation

End Goal

Complete AI-powered recruitment solution

Thank you for your attention and engagement throughout this presentation.

— Ravichandra D | Infosys Springboard Internship 6.0

This AI Skill Gap Analyzer project represents a significant step toward revolutionizing recruitment through intelligent document processing. Your feedback and questions are welcome as we continue to enhance this system for real-world applications.