Resume Analyzer Documentation

Overview

The Resume Analyzer is a powerful Python-based tool designed to streamline the processing, analysis, and extraction of insights from resumes. By leveraging Google's Gemini AI (Gemini 1.5-flash) and Google Drive API, this tool automates the extraction of detailed information such as candidate skills, education, and AI/ML expertise. The results are stored in structured Excel files, making it easier for recruiters to evaluate candidates effectively.

Features

- **AI-Powered Analysis**: Utilizes Gemini 1.5-flash for accurate resume parsing.

- **Google Drive Integration**: Automates the download of resumes from shared Google Drive folders.

- **Batch Processing**: Handles concurrent processing of multiple resumes for scalability.

- **Comprehensive Scoring**: Evaluates candidates on AI/ML and Gen AI experience.

- **Detailed Reporting**: Generates Excel reports with multiple sheets for in-depth analysis.

Technical Architecture

Core Components

1. **ResumeAnalyzer Class**

   - Handles all processing logic, including Google Drive and Gemini AI integrations.

   - Implements concurrent processing and robust error handling.

2. **Integration Services**

   - **Google Drive API**: Automates file management.

   - **Gemini AI**: Extracts structured insights from unstructured resume data.

   - **PyPDF2**: Handles text extraction from PDF files.

3. **Data Processing Pipeline**

PDF Files → Text Extraction → AI Analysis → Data Aggregation → Excel Report

---

Implementation Details

Libraries Used

- **Generative AI SDK**: Integration with Gemini 1.5-flash model.

- **PyPDF2**: Text extraction from PDFs.

- **Google Drive API**: Access and download files from Google Drive.

- **pandas**: Data organization and processing.

- **concurrent.futures**: Enables parallel processing of resumes.

- **tqdm**: Tracks batch processing progress.

- **openpyxl**: Creates multi-sheet Excel reports.

Workflow Steps

1. **Setup**:

   o Configure the Gemini AI API key and initialize the Gemini model.

   o Authenticate Google Drive API using a service account JSON file.

2. **Resume Processing**:

   o List and download resumes from a specified Google Drive folder.

   o Extract text from PDF files using PyPDF2.

   o Analyze resumes using Gemini AI with custom prompts.

3. **Batch Processing**:

   o Process resumes in batches of 10 for optimal performance.

   o Log errors and handle failures gracefully.

4. **Excel Output**:

   o Generate a summary sheet with key insights.

   o Include detailed sheets for projects, skills, and failed files.

Key Features and Functionality

1. Generative AI Integration

- **Natural Language Understanding**:

  o Extracts structured data such as name, education, and skills.

  o Employs few-shot examples for improved context comprehension.

- **Scoring System**:

  o Evaluates AI/ML experience and generative AI expertise based on predefined rules.

2. Google Drive Integration

- Automates the download of resumes stored in specified Google Drive folders.

- Ensures seamless integration with shared folders using service account credentials.

3. Innovative Scoring System

- **Gen AI Experience (1-3 scale)**:

    o Score 1: Basic exposure.

    o Score 2: Hands-on projects.

    o Score 3: Advanced implementation.

- **AI/ML Experience (1-3 scale)**:

    o Score 1: Basic knowledge.

    o Score 2: Project experience.

    o Score 3: Advanced expertise.

4. Comprehensive Excel Reporting

- **Resume Analysis**: Contains core candidate information such as name, education, and scores.

- **Projects**: Details candidate projects, including descriptions and technologies used.

- **Skills Analysis**: Provides a breakdown of skills and scores.

- **Failed Files**: Lists problematic files with error messages.

Setup Instructions

Prerequisites

- Python 3.7+, Jupyter Notebook, Notepad and Command prompt

- Required Python packages:

    - !pip install PyPDF2
    - !pip install pandas
    - !pip install google-generativeai
    - !pip install --upgrade google-generativeai
    - !pip install tqdm
    - !pip install google-api-python-client
    - !pip install google-auth

- Google API credentials (service account JSON file).

- Gemini AI API key.

Configuration

1. **Google Drive Setup**:
   - Create a service account in Google Cloud Console.
   - Download the credentials JSON file.
   - Share the target Drive folder with the service account email.
2. **Gemini AI Setup**:
   - Obtain the Gemini API key.
   - Configure it in the code initialization.

Steps to Run the Code

Installation:

- **Python:** Make sure you have Python installed. It's recommended to use a Python distribution like Anaconda, which comes with many scientific computing packages and Jupyter pre-installed. You can download it from https://www.anaconda.com/.

- **Jupyter:** If you don't use Anaconda, you can install Jupyter using pip:

bash - pip install jupyter

Starting Jupyter Notebook:

- Open your terminal or command prompt.
- Navigate to the directory where you want to create your notebook files using the cd command (e.g., cd Documents/MyNotebooks).
- Type jupyter notebook and press Enter.

Using Jupyter Notebook:

- If you have already save the code in your system.
- Retrieve the path or find the code file where it is and open it. And run the code.
- Click the "New" button in the top right corner and select "Python 3" (or your desired kernel). This will create a new notebook file.
- The notebook consists of cells. You can write code in code cells and text (using Markdown) in Markdown cells.
- To run a code cell, press Shift + Enter or click the "Run" button in the toolbar. The output of the code will be displayed below the cell.

Output File: Excel Format

Excel File Structure

1. **Resume Analysis**:

   o Name, contact details, university, year of study, course, discipline, key skills, CGPA, key skills, Gen AI score, AI/ML score, supporting certificates, career potential, projects, technical insights, experience level.

2. **Projects**:

   o Candidate name, project name, description, technologies used.

3. **Skills Analysis**:

   o Name, skills, AI/ML score, Gen AI score.

4. **Failed Files**:

   o List of resumes that failed to process with error messages.

Error Handling and Reliability

- Retry mechanism for API calls.

- Concurrent processing with error isolation.

- Comprehensive logging for debugging.

- Tracks and reports failed files for further review.

Limitations and Future Enhancements

Current Limitations

- Limited support for non-English resumes.

- Heavily reliant on PDF text extraction quality.

Future Enhancements

1. **Multilingual Support**:

   o Incorporate language detection and translation for non-English resumes.

2. **Customizable Scoring Criteria**:

   o Allow users to define their own scoring rules.

3. **Enhanced Error Handling**:

   o Improve resilience against poorly formatted resumes.

4. **Support for Additional File Formats**:

   o Extend support to DOCX and other document types.

5. **Real-Time Processing**:

o   Enable real-time resume analysis.

Conclusion

The Resume Analyzer is a cutting-edge tool that automates and enhances the resume analysis process using Generative AI. Its ability to process large datasets efficiently, provide insightful scoring, and generate detailed reports makes it an invaluable resource for recruiters. Planned enhancements will further strengthen its capabilities, ensuring scalability, flexibility, and ease of use.