

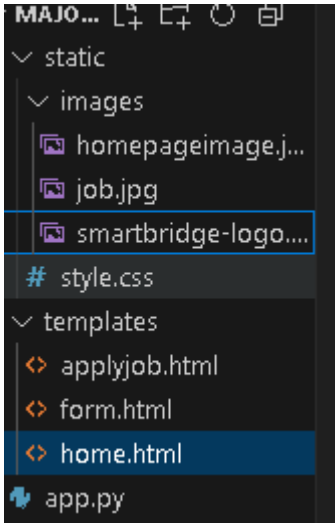
Data Collection and Preprocessing Phase

Date	Nov 2024
Team ID	Team-739663
Project Title	AI-Enabled Candidate Resume Screening using NLP
Maximum Marks	6 Marks

PreparationTemplate

The data collection and preparation phase for the **AI-enabled candidate resume screening project using NLP** involves several key steps. Firstly, resumes and job descriptions are collected from various sources, including company HR databases, publicly available resume datasets, and job portals. The collected documents, which are often in different formats (PDF, DOCX, TXT), are then preprocessed to ensure consistency and structure.

Section	Description
Data Overview	Collect resume and job description data from multiple sources, including: – Company HR databases and applicant tracking systems (ATS) – Public resume datasets (e.g., Kaggle) – Job portals and career websites.
Data Preparation	The preparation phase for the AI-enabled resume screening project involves: – Parsing resumes and job descriptions – Text normalization (lowercasing, punctuation removal) – Tokenization and stop word removal.
Handling missing values	Address missing skills, education, or experience data using imputation techniques such as: – Statistical methods (mean/mode for structured fields) – Rule-based inference or ML-based completion for unstructured fields.

Handling Outliers in Data	<p>Remove or transform outliers in attributes like years of experience or education levels using:
 – Data trimming
 – Log or z-score transformations
 – Robust regression for profile score prediction.</p>
<h2>Install Rasa and Dependencies and Data Preparation</h2>	
Installation of Rasa	<p>Once the virtual environment is active, install Rasa and its dependencies using:
</p> <pre>>>> pip install rasa</pre>
Setting up Rasa Project	<p>Initialize a new Rasa project using:
</p> <pre>>>> rasa init</pre>
Collecting the data: file	<p>The file is used to define training examples for understanding user inputs like uploading a resume, requesting job matches, or querying skills.</p> 

home.html

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>SmartBridge - About Us</title>
    <link rel="stylesheet" href="{{ url_for('static', filename='style.css')}}">
  </head>
  <body>
    <header class="navbar">
      <div class="logo">
        
      </div>

      <nav class="nav-links">
        <a href="/applyjob">
          <button>Apply job </button>
        </a>
      </nav>
    </header>
  </body>
</html>
```

applyjob.html

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>SmartBridge - Job Prediction</title>
    <link rel="stylesheet" href="{{ url_for('static', filename='style.css')}}">
  </head>
  <body>
    <header class="navbar">
      <div class="logo">
        
      </div>

      <nav class="nav-links">
        <a href="/form">
          <button>Apply Job</button>
        </a>
      </nav>
    </header>
  </body>
</html>
```