**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 9 JULY 2024 |
| Team ID | INDIVIDUAL |
| Project Title | CareerCraft: ATS-Optimized Resume Analyzer using Gemini Model |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | CareerCraft optimizes resumes for ATS, helping job seekers match their resumes with job descriptions to improve interview chances. |
| Data Collection Plan | Data will be collected from:   * **User-uploaded Resumes:** PDF or DOCX formats from users. * **Job Descriptions:** Sourced from online job portals. |
| Raw Data Sources Identified |  **User Resumes:** Resumes uploaded by users in PDF or DOCX formats, stored locally with private access.   **Job Descriptions:** Text data collected from various online job portals, publicly accessible through web URLs. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| User Resumes | Resumes uploaded by users | Local storage | PDF, DOCX | Varies | Private |
| Job Descriptions | Job listings collected from online portals | Web URLs | Text | Varies | Public |