**Report**

**Approach:**

1. **PDF Data Extraction:**
   * PDF text extraction was performed using the PyMuPDF (Fitz) library.
   * Text data from CVs was extracted from a folder structure where CVs were organized by department.
   * Key details extracted included Category (Job role), Skills, and Education.
2. **Job Description Data Understanding:**
   * Job descriptions were loaded from a CSV file containing columns like 'job\_description,' 'company\_name,' 'position\_title,' 'description\_length,' and 'model\_response.'
3. **Candidate-Job Matching:**
   * DistilBERT was used for tokenization and embedding extraction of both job descriptions and CV details.
   * Cosine similarity was calculated between job descriptions and CVs to measure similarity.
   * The top 5 candidates for each job description were ranked based on the highest similarity scores.

**Challenges Faced and Solutions:**

1. **PDF Extraction:**
   * Challenge: Variability in PDF formats and structures.
   * Solution: Utilized PyMuPDF (Fitz) for text extraction, which provided good results across different PDF layouts.
2. **Data Preprocessing:**
   * Challenge: Standardizing extracted CV data.
   * Solution: Implemented data preprocessing techniques to extract key details, such as skills and education, from unstructured text.
3. **Matching Algorithm:**
   * Challenge: Matching CVs with job descriptions effectively.
   * Solution: Utilized cosine similarity as a measure to match CVs with job descriptions based on their semantic similarity.