

Report Title: Matching Job Descriptions with CVs

1. Approach to the Task:

In this project, my primary objective was to match job descriptions with candidate CVs based on their content. I followed a systematic approach that involved the following steps:

Data Collection: I collected job descriptions and candidate CVs from respective sources.

Text Preprocessing: I cleaned and tokenized the text data, after that performed text embedding.

Embedding Extraction: I utilized a pretrained model (DistilBERT) from the Transformers library to convert text data into embeddings, allowing us to capture the semantic meaning.

Similarity Calculation: I calculated the cosine similarity between job descriptions and CVs to determine the degree of match.

Candidate Ranking: Based on similarity scores, I ranked candidates for each job description.

CSV Data Handling: I handled CSV files for input and output.

2. Challenges Faced and Solutions:

Semantic Matching: Cosine similarity, while effective, may not capture semantic similarities accurately. To address this, advanced techniques such as Word2Vec or Doc2Vec can be used for improved semantic matching.

Data Quality: Ensuring data quality through preprocessing, including removing special characters and stopwords, was crucial for accurate results.

Resource Management: Efficient resource management is essential, especially when working with large datasets or computationally intensive models.

3. Top 5 Candidates for Each Job Description:

Based on the similarity scores, I identified the top 5 candidates for each job description. These candidates are considered the best matches for the respective roles. We only used first 15 job descriptions.

Job 1

INFORMATION-TECHNOLOGY
CONSULTANT
CONSULTANT
CONSULTANT
ARTS

Job 2

APPAREL
ARTS
INFORMATION-TECHNOLOGY
ARTS
SALES

Job 3

APPAREL
ARTS
CONSULTANT
SALES
BANKING

Job 4

DIGITAL-MEDIA
HR
ARTS
HEALTHCARE
TEACHER

Job 5

INFORMATION-TECHNOLOGY
APPAREL
DIGITAL-MEDIA
ARTS
INFORMATION-TECHNOLOGY

Job 6

DESIGNER
FITNESS
DESIGNER
INFORMATION-TECHNOLOGY
ARTS

Job 7

DIGITAL-MEDIA
INFORMATION-TECHNOLOGY
DESIGNER
CONSULTANT
HR

Job 8

ARTS
ADVOCATE
CONSULTANT
INFORMATION-TECHNOLOGY
HEALTHCARE

Job 9

DIGITAL-MEDIA
APPAREL
CONSULTANT
HR
DESIGNER

Job 10

DIGITAL-MEDIA
CONSULTANT
TEACHER
INFORMATION-TECHNOLOGY
AVIATION

Job 11

DIGITAL-MEDIA
HR
CONSULTANT
SALES
INFORMATION-TECHNOLOGY

Job 12

APPAREL
INFORMATION-TECHNOLOGY
CONSULTANT
ARTS
BANKING

Job 13

ARTS
APPAREL
DIGITAL-MEDIA
INFORMATION-TECHNOLOGY
DESIGNER

Job 14

HEALTHCARE
AGRICULTURE
INFORMATION-TECHNOLOGY
AUTOMOBILE
AVIATION

Job 15

ARTS
APPAREL
DESIGNER
DIGITAL-MEDIA
BANKING

4. Recommendations and Insights:

Enhance Skill Matching: Consider using named entity recognition (NER) to identify specific skills or technologies mentioned in job descriptions and resumes for more accurate skill matching.

Include Additional Features: Incorporate more features like location, experience level, or industry to make the matching process more comprehensive.

Scale for Larger Datasets: For larger datasets, explore distributed computing frameworks like Apache Spark.

Effective Use of Pretrained Models: Pretrained models like DistilBERT provided efficient text embedding, reducing the need for manual feature engineering.