Data Requirements for JD-CV Matching Research

1. Objective

To effectively evaluate and improve the **job description (JD) to CV matching process**, we require access to a structured dataset that reflects real-world hiring scenarios within the company. This will allow us to test different matching algorithms and measure their performance accurately.

2. Required Data

2.1 Job Descriptions (JDs)

• Number of available JDs: [Estimate]

• Data points needed per JD:

  • **Job Title**

    • **Company Description**

    • **Job Description (Full Text)**

    • **Required & Preferred Skills**

    • **Experience Level (Years)**

    • **Education Requirements**

    • **Industry / Department - if Affinda derives it**

    • **Job Posting Date**

2.2 Resumes / CVs

• Number of available CVs: [Estimate]

• Data points needed per CV:

    • **Candidate Name (or Anonymized ID)**

    • **Candidate Summary**

    • **Work Experience (Job Titles, Companies, Years)**

    • **Volunteering**

    • **Projects (consulting, etc.)**

    • **Education History**

    • **Skills & Certifications**

    • **Languages Spoken**

    • **Keywords or Tags (if available)**

2.3 Existing Matching Data

• **Historical matches (successful JD-CV pairs)**

   • Number of available matches: ~100

   • Criteria used to determine a "successful" match

• **Unsuccessful applications (if available)**

   • Allows for understanding false positives/negatives in past matches

3. Data Format & Access

• Preferred format: **CSV, JSON, or database access**

• If possible: **Affinda access**

• Any existing **structured or labeled datasets** would be helpful

• Expected delivery: **[Propose a date based on project timeline]**

4. Data Privacy & Compliance Considerations

• Ensure that all data shared complies with **GDPR and company policies**.

• If sensitive information is present, **anonymization or pseudonymization** would be preferred.

• Please confirm if any **restrictions** apply to external processing or model testing.

5. Next Steps

• Let us know the **feasibility** of providing this data and any limitations.

• If needed, we can schedule a short call to clarify requirements.