Machine Learning Semester Project Proposal

Project Title: Resume-Job Match Scorer Using NLP and Semantic Similarity

Team Members:

1. A.B. Ghalib - BSDSF22M047

2. Zohha Azhar - BSDSF22M042

3. Laiba Arshad - BSDSF22M031

1. Problem Statement

Job seekers often struggle to understand how well their resume aligns with specific job descriptions. Recruiters face a similar challenge in quickly screening resumes. A tool that automatically scores and explains the relevance between a resume and a job posting can save time, improve applicant-job fit, and reduce hiring mismatches.

2. Objectives

- To develop an ML-based system that measures the semantic similarity between a resume and a job description.
- To provide a numeric match score and actionable feedback (e.g., missing skills, alignment issues).
- To deploy a simple web interface for user interaction.

3. Proposed Methodology

- Text Preprocessing: Clean text (remove stopwords, punctuation, etc.), tokenize.
- Feature Extraction: Use TF-IDF vectors and cosine similarity for a baseline.
- Advanced Embeddings: Use BERT or Sentence-BERT for deep semantic comparison.
- Scoring System: Generate a match score and highlight key similarities/differences.
- Feedback: Extract skill/keyword gaps between resume and job description.
- Web App: Flask or Streamlit-based UI to upload text files and show scores/feedback.

4. Dataset Description

We will use publicly available resume-job description datasets such as the CareerBuilder dataset, or collect sample job-resume pairs from platforms like Kaggle and Indeed. These will be augmented

with synthetic/resampled data for fine-tuning.

5. Expected Outcomes

- A functional model that gives a resume-job similarity score using NLP techniques.

- A web-based interface where users upload documents and get matching feedback.

- Insights into key gaps between resumes and job postings for career improvement.

6. Timeline (3-4 Weeks)

- Week 1: Finalize dataset; perform preprocessing and baseline similarity tests

- Week 2: Implement BERT-based embedding models; evaluate scoring methods

- Week 3: Develop and test feedback/explanation system; build web app frontend

- Week 4: Final testing, polishing UI, prepare final report and presentation

7. Originality & Feasibility

This project applies practical NLP methods to a real-world HR problem. Using pre-trained language models like BERT allows us to focus on fine-tuning and integration rather than building from scratch. The 3-4 week timeline is realistic given the modular scope.