# Project Proposal

## Title

**Resume Screening and Job Matching System using Natural Language Processing**

---

## 1. Introduction

In today's recruitment process, organizations receive a large number of resumes for each job opening. Manually screening these resumes is time-consuming, error-prone, and inefficient. With the growth of digital recruitment platforms, there is a strong need for automated systems that can analyze unstructured resume data and match candidates to job requirements effectively.

This project proposes the development of a **Resume Screening and Job Matching System** using **Natural Language Processing (NLP)** and **Machine Learning (ML)** techniques. The system will automatically extract text from resumes, process unstructured textual data, and compute a relevance score between resumes and job descriptions to assist recruiters in shortlisting suitable candidates.

---

## 2. Problem Statement

Manual resume screening is inefficient and often inconsistent due to human bias and time constraints. Recruiters need an automated system that can: - Process resumes in PDF format - Understand unstructured textual data - Match resumes with job descriptions - Rank candidates based on relevance

The lack of such automated tools leads to delayed hiring decisions and missed suitable candidates.

---

## 3. Objectives

The main objectives of this project are: 1. To extract text from PDF resumes automatically. 2. To preprocess and clean unstructured resume and job description text. 3. To build an NLP-based model using TF-IDF and Machine Learning for resume classification. 4. To implement semantic similarity scoring using BERT embeddings. 5. To rank resumes based on their relevance to a given job description. 6. To deploy the system as a simple API or web application.

---

## 4. Scope of the Project

The scope of this project includes: - Resume parsing from PDF documents - Text preprocessing and feature extraction - Resume–job matching using similarity metrics - Candidate ranking - Deployment of the system for practical usage

The project does not aim to replace human recruiters but to assist them by automating the initial screening process.

---

## 5. Proposed System Architecture

The system will consist of the following modules:

1. **Resume Parser**: Extracts text from uploaded PDF resumes.
2. **Text Preprocessing Module**: Cleans and normalizes text data.
3. **Feature Extraction Module**: Generates TF-IDF vectors and BERT embeddings.
4. **Matching & Scoring Module**: Computes similarity scores using cosine similarity.
5. **Ranking Module**: Ranks resumes based on match scores.
6. **Deployment Layer**: Provides access through a web interface or REST API.

---

## 6. Methodology

### 6.1 Data Collection

- Publicly available resumes and job descriptions
- Synthetic resumes created for academic purposes

### 6.2 Text Preprocessing

- Lowercasing
- Removal of punctuation and stopwords
- Lemmatization

### 6.3 Feature Engineering

- TF-IDF vectorization for baseline modeling
- BERT-based sentence embeddings for semantic similarity

### 6.4 Modeling

- Machine Learning classifiers (Logistic Regression / SVM)
- Cosine similarity for resume-job matching

### 6.5 Evaluation

- Precision, Recall, and F1-score for classification
- Ranking effectiveness using top-k relevance

---

## 7. Tools and Technologies

- **Programming Language**: Python
- **Libraries**: Pandas, NumPy, Scikit-learn, NLTK / spaCy
- **Deep Learning**: BERT (Sentence Transformers)
- **Resume Parsing**: pdfplumber / PyMuPDF
- **Deployment**: FastAPI / Streamlit
- **Version Control**: Git and GitHub

---

## 8. Expected Outcomes

- An automated resume screening system
- Improved efficiency in candidate shortlisting
- Accurate resume–job relevance scoring
- Practical exposure to NLP and ML techniques

---

## 9. Applications

- Recruitment platforms
- Applicant Tracking Systems (ATS)
- HR analytics tools
- Campus placement automation

---

## 10. Project Timeline

| Phase | Duration |
|---|---|
| Requirement Analysis | 1 Week |
| Data Collection & Preprocessing | 1 Week |
| Model Development | 2 Weeks |
| Evaluation & Testing | 1 Week |
| Deployment & Documentation | 1 Week |

---

## 11. Conclusion

The proposed Resume Screening and Job Matching System aims to automate and enhance the recruitment process using NLP and Machine Learning. By efficiently handling unstructured resume data and providing accurate candidate ranking, the system can significantly reduce manual effort and improve hiring decisions.

This project provides strong academic as well as industry relevance and serves as a practical application of data science concepts.

---

## 12. References


1. Scikit-learn Documentation
2. Hugging Face BERT Documentation
3. Research papers on resume screening and NLP-based matching