ITSOLERA Pvt. Ltd.

Machine Learning Internship Project Proposal

Project Title: Intelligent Resume Screening System Using Machine Learning

1. Introduction

In recruitment, companies receive many resumes for a single job position. Checking each resume manually takes a lot of time and effort. This project focuses on building an **Intelligent Resume Screening System** that helps recruiters automatically shortlist candidates.The system uses **Natural Language Processing (NLP)** to analyze resumes and a job description, then ranks resumes based on how well they match the job requirements.

2. Objectives

The objectives of this project are:

- To automate resume shortlisting
- To reduce manual screening effort
- To understand basic NLP techniques
- To rank resumes based on job relevance

3. Literature Review

Many studies have explored the use of **Natural Language Processing (NLP)** and **Machine Learning** in recruitment systems. Resume screening is a common application where text-based analysis is used to match candidate profiles with job requirements.Techniques such as **text preprocessing, TF-IDF vectorization, and cosine similarity** are widely used to measure similarity between resumes and job descriptions. These approaches help automate the hiring process, reduce human effort, and improve shortlisting accuracy. Similarity-based ranking systems have shown effective results in identifying relevant candidates for specific job roles.

4. Methodology

4.1. Data Collection

We will use a set of **sample resumes** collected in text or PDF format along with one or more **job descriptions**. These resumes represent different candidate profiles with varying skills and experience levels.

4.2. Data Preprocessing

Text Cleaning: Convert text to lowercase and remove punctuation and special characters.
- **Stopword Removal:** Remove commonly used words that do not add meaningful information.
- **Tokenization:** Split text into individual words.
- **Lemmatization:** Convert words into their base form to improve matching accuracy.

4.3. Feature Extraction and Similarity Calculation

- **Text Vectorization:** Use TF-IDF Vectorizer to convert resumes and job descriptions into numerical vectors.
- **Similarity Measurement:** Apply Cosine Similarity to calculate how closely each resume matches the job description.
- **Ranking:** Rank resumes based on similarity scores from highest to lowest.

4.4. Model Evaluation

- **Similarity Score Analysis:** Evaluate how well resumes match the job description.
- **Ranking Accuracy:** Check whether relevant resumes appear at the top of the ranking list.
- **Manual Validation:** Compare system rankings with expected results for validation.

5. Expected Outcomes

- An automated resume screening system based on NLP techniques.
- Ranked list of resumes according to job relevance.
- Reduced manual effort in resume shortlisting.
- Improved understanding of practical NLP applications in recruitment systems.

6. Conclusion

This project focuses on developing an intelligent resume screening system using Machine Learning and Natural Language Processing techniques. By analyzing resumes and job descriptions, the system helps recruiters shortlist candidates efficiently. The project provides students with hands-on experience in NLP, similarity analysis, and real-world HR technology applications.