**Title: Automated Resume Matching and Ranking System Using NLP and Neural Networks**

Date: 3rd October, 2024

**Objective**

The primary objective of this project is to develop an automated solution to match resumes with job descriptions and rank them based on their relevance. The aim is to streamline the hiring process by providing an efficient way to score and rank candidates based on the requirements of specific job openings.

**Problem Statement**

Manual resume screening is time-consuming and prone to human bias. Recruiters often need to filter through numerous resumes to identify the most qualified candidates for specific job roles. The key challenge is to match candidates' qualifications, experience, and skills against job descriptions accurately and efficiently.

**Approach and Methodology**

* **Resume Extraction:** Extracting Meaningful information from resume
  + **Text Extraction:** Collect Candidate Details: (e.g., **Name**, **Mobile number**, **Email**, **College Name,** **Skills**, **Experience**, **Company Names**, **Number of years Experience,** **Number of Pages resume**, and other text) from resumes using Natural Language Processing (NLP) and regex from PDF or DOCX.
  + **Pre-Process:** 
    - **Text Cleaning:** Remove URLs, Remove Social Media Indicators, Remove Hashtags, Remove Punctuation, Remove Non-ASCII Characters, Remove Extra Whitespace
    - **Lowercasing**
    - **Stopwords Removal**
* **Job\_Description:** Extracting Meaningful information from resume
  + **Text Extraction:** Collect Job Description from PDF or DOCX.
  + **Pre-Process:** 
    - **Text Cleaning:** Remove URLs, Remove Social Media Indicators, Remove Hashtags, Remove Punctuation, Remove Non-ASCII Characters, Remove Extra Whitespace
    - **Lowercasing**
    - **Stopwords Removal**
* **Resume Multi Classification:** Predicting suitable job role from a resume
  + **Data Loading:** A CSV file containing resumes and their respective categories is loaded.
  + **Pre-Process: Data Cleaning**
    - **Text Cleaning:** Remove URLs, Remove Social Media Indicators, Remove Hashtags, Remove Punctuation, Remove Non-ASCII Characters, Remove Extra Whitespace
    - **Lowercasing**
    - **Tokenization**
    - **Stopwords Removal**
    - **Data Encoding**
    - **Data Splitting**
    - **Tokenization and Padding**
  + **Building the Neural Network Model:** Bidirectional LSTM Model v/s Global Max Pooling Model
  + **Model Compilation and Training**
  + **Evaluation Metrics:** 
    - **Accuracy:** The **Bidirectional LSTM model** achieved an **accuracy of 99%** on the test data, while the Global Max Pooling model **achieved 98%.**
  + **Making Predictions: A dictionary of 5 Probabilitie of roles will get based on ascending order.**
* **Resume vs Job Description Comparison** 
  + **Embeddings :** Used **Hugging Face Transformers: BERT**(sentence-transformers/all-mpnet-base-v2) It is a variant of BERT specifically designed for generating sentence embeddings, optimizing the model for tasks like semantic textual similarity
  + **Matching:** Cosine similarity