***Topic: NLP-Based Resume Screening Application***

**Abstract:**

The objective of this project is to develop an NLP-based resume screening application that automates the process of categorizing resumes and extracting key information from them. This tool utilizes Natural Language Processing (NLP) techniques to identify the category of the resume, extract candidate details, and recognize relevant skills and technologies. By leveraging machine learning models, the application enhances the efficiency and accuracy of the recruitment process. A comprehensive dataset of resumes labelled by job categories is used to train the model, which is then deployed in a user-friendly web application. The final deliverable is a streamlined resume screening system that supports recruiters in making informed hiring decisions.

**Dataset and Problem Statement:**

The dataset comprises a collection of resumes, each labelled with a job category. The resumes are in textual format and contain various sections such as contact information, work experience, education, skills, and certifications. The primary challenge is to accurately classify these resumes into predefined job categories and extract essential information like candidate name, technologies known, and skills.

**Approach:**

1. **Data Preprocessing:** Clean the resume text to remove unwanted characters and standardize the format.
2. **Feature Extraction:** Use Term Frequency-Inverse Document Frequency (TF-IDF) to transform the textual data into numerical features.
3. **Model Training:** Train a classification model using the pre-processed dataset. The model will learn to categorize resumes into job categories based on the features extracted.
4. **Named Entity Recognition (NER):** Implement NER techniques to extract key details such as candidate names and known technologies from the resumes.
5. **Application Development:** Develop a web application using Stream lit to allow users to upload resumes, view the predicted job category, and extract key information.

**Solution:**

We solve the problem by developing an end-to-end system that integrates data preprocessing, model training, and a user interface for resume screening. The solution employs machine learning models to categorize resumes and NLP techniques to extract information. Using the nltk library for preprocessing and scikit-learn for model training ensures robust performance. The system offers a user-friendly interface for easy interaction.

**Why NLP:**

NLP is crucial in this application for understanding and processing human language in resumes. It allows the system to handle unstructured data, extract meaningful insights, and make accurate predictions. By applying NLP, we can automate the resume screening process, reduce manual effort, and improve the accuracy of candidate selection.

**Application (app.py):**

The application loads pre-trained models (clf for classification and tfidf for feature extraction), processes uploaded resumes to clean the text, and transforms it into features. The cleaned resume is then classified into a job category using the loaded model. The application also extracts candidate names and known technologies, and generates job search links for the predicted category. It displays the predicted category, candidate name, and technologies known, and provides buttons for job search on popular job portals like Indeed, Glassdoor, and Monster.

This approach ensures an efficient and automated system for resume screening, leveraging the power of NLP and machine learning to assist recruiters in the hiring process.

**Reference:**

* Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc.