NAME: Muzammil Amjad

Project Title: Resume Screening System Using NLP

Description:
This project aims to develop an automated resume screening system using Natural Language Processing (NLP) and machine learning algorithms. The system processes resume text data through advanced text preprocessing and vectorization techniques, such as TF-IDF or Word Embeddings, to extract relevant features like skills, experience, and education. It then applies classification models to match resumes to job descriptions, ranking candidates based on their suitability. The system supports recruiters by efficiently filtering large volumes of resumes, reducing manual effort and improving hiring accuracy.

Goal:
To classify and rank resumes based on their relevance to job descriptions using NLP techniques and machine learning models, including K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machines (SVM), and boosting algorithms (XGBoost, LightGBM).

Dataset Link: https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset/code

Tech Stack:
Programming Language: Python
Frontend: Web App using NGROK, HTML rendering
Libraries pandas, numpy, sklearn, nltk, spacy, xgboost, lightgbm, matplotlib, seaborn,
Model: K-Nearest Neighbors, Logistic Regression, Support Vector Machines, XGBoost, LightGBM

Components:

- Text preprocessing (cleaning, tokenization, stopword removal, lemmatization)
- Feature extraction using TF-IDF or pre-trained Word Embeddings (e.g., Word2Vec, GloVe)
- Model training and evaluation (accuracy, precision, recall, F1-score, confusion matrix)
- Model comparison between different classifiers for performance optimization