Personal Project Report

Project Title: Resume Screening App using LLM (Large Language Models)

Brief Description:

The Resume Screening App is a machine learning-powered web application designed to automate and streamline the process of analyzing and categorizing resumes. It uses Natural Language Processing (NLP) techniques and machine learning algorithms to score resumes, predict relevant job categories, and provide insightful feedback to improve resume quality. The tool enhances the recruitment process by providing fast, consistent, and intelligent resume analysis through an interactive interface.

Deployed Live App Link: HuggingFace Space - Resume Screening App

GitHub Repo Link: Resume Screening App

Objectives:

- Automate resume parsing and skill extraction using NLP.
- Score resumes based on their relevance to predefined job categories.
- Classify resumes into various job roles using ML algorithms.
- Visualize submission statistics using interactive charts.
- Provide actionable feedback to help candidates improve their resumes.

Tools/Tech Used:

- **Programming Language:** Python
- **Libraries:** NLTK, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn
- **Vectorization:** TF-IDF (Term Frequency-Inverse Document Frequency)
- **ML Model:** K-Nearest Neighbors (KNN) with OneVsRestClassifier
- **Web Framework:** Flask
- **Environment:** Jupyter Notebook, VS Code

- **Deployment:** Hugging Face Spaces

Methodology/Implementation:

1. Data Preprocessing:

- Loaded a labeled dataset (UpdatedResumeDataSet.csv) with 962 resumes.
- Applied regex-based cleaning to remove URLs, special characters, non-ASCII text, and excessive spacing.
- Encoded the target labels (job categories) using Label Encoding.

2. Visualization & Exploration:

- Created bar and pie charts to visualize category distribution.
- Explored resume content using word frequency and category alignment.

3. Feature Engineering:

- Transformed resumes into numerical vectors using TF-IDF.
- Captured essential skills and keywords from resumes.

4. Model Training:

- Split the dataset into training and testing sets (80/20).
- Trained a KNeighborsClassifier within a OneVsRestClassifier wrapper.
- Achieved **98.4% accuracy** on the test set.

5. Prediction System:

- Enabled resume category prediction based on a new resume input.
- Cleaned the input, vectorized it using the saved TF-IDF model, and predicted using the trained classifier.
- Mapped predicted labels back to job category names.

6. Deployment:

- Pickled the trained TF-IDF vectorizer and ML model (tfidf.pkl, clf.pkl).
- Deployed a Flask-based interface on Hugging Face for public access.

Validation:

- Manually tested with multiple resume inputs to ensure proper cleaning and categorization.
- Verified category accuracy using known examples.
- Evaluated model using accuracy_score, yielding high accuracy.
- Ensured deployment was stable and responsive for real-time use.

Results/Performance:

- Successfully predicted job roles from resume text with high precision.
- Supported resume uploads in formats like PDF, DOCX, and TXT.
- Provided users with resume scores, relevant job matches, and suggestions for improvement.
- Simplified the screening process for recruiters and job seekers.

Future Work:

- Integration with LLMs (like GPT) for deeper skill extraction and semantic analysis.
- Enhanced feedback system using GPT-generated suggestions.
- Dashboard analytics for bulk resume processing.
- Role-specific resume tailoring using job description comparison.
- Expand dataset with more diverse roles and real-world resumes.

Conclusion:

The **Resume Screening App using LLM** effectively demonstrates the synergy between machine learning, NLP, and web technologies to solve a real-world HR challenge. By automating resume classification and providing intelligent feedback, the app improves the efficiency and accuracy of the hiring process. With high model performance and a user-friendly interface, this project lays the groundwork for future AI-powered HR solutions, offering scalable tools for recruiters and meaningful feedback for job seekers.