Resume Categorization

This project categorizes resumes into respective domain categories (e.g., sales, marketing) using a machine learning model. The project involves training a model, categorizing resumes, and storing the results in both categorized folders and a CSV file.

Prerequisites

- **Python 3.x**: Make sure Python is installed on your system.
- Install required Python packages by running: ```bash pip install -r requirements.txt

Directory Structure

- **dataset/data/data/**: Contains the resumes arranged in category folders.
- **dataset/task_result/models/**: Contains the resumes arranged in category folders.
- **dataset/task_result/categorized_resumes/**: Contains the resumes arranged in category folders.
- **dataset/task_result/csv_resume/**: Contains the resumes arranged in category folders.

How to Use

1. Train the Model

- Ensure resumes are placed in dataset/data/data/, categorized into subdirectories.
- Split the data, preprocess, and train the model.
- Save the trained model and vectorizer in dataset/new/models/.

2. Run the Categorization Script

- Run the categorization script with the following command: ```bash python script.py dataset/data/data

3. Output

- Categorized resumes will be copied into 'dataset/task_result/categorized_resumes/.'
- The CSV file containing the categorization results will be stored in 'dataset/task_result/csv_resume/.'

Notes

- Original resumes remain in their original directories.
- The categorization results are saved in a CSV file

Project Directory Structure

The following is the structure of the project directory for the Resume Categorization Task.

```
resume-categorization-ml/
├── dataset/
│   ├── data/
│   │   └── data/               # Original resumes categorized into folders (e.g., ACCOUNTANT, MARKETING)
│   └── Resume/                 # Resume.csv categorized resume in .csv file
│   └── task_result/
│       ├── models/             # Trained model and vectorizer
│       ├── categorized_resumes/  # Categorized resumes in folders, organized by predicted category
│       └── csv_resume/
│           └── categorized_resumes.csv  # CSV file with category columns
├── eda.ipynb                   # Script for EDA dataset available
├── model_generation.ipynb      # Script for model training and validation
├── script.py                   # Script for categorizing resumes
├── requirements.txt            # List of required Python packages
├── README.md                   # Readme file with project overview and instructions
```