Resume Analysis Tool

DATS 6312: Natural Language Processing

Fall 2023

December 11th, 2023

By

Jahnavi Ramagiri & Shubham Jadhav

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Introduction

Objective:

- Elevate Recruitment efficiency through elaborate Resume Analysis

Key Features:

- Job Role Matching:
  - Multilabel Classification of Resumes to match relevant Job roles
- Smart Resume Analysis:
  - Resume Summarisation
  - Resume Skill Extraction
- Resume - Job Description Similarity
  - Similarity Score for resume-job description
- Custom Criteria:
  - Tailor evaluations to match specific confidence thresholds.

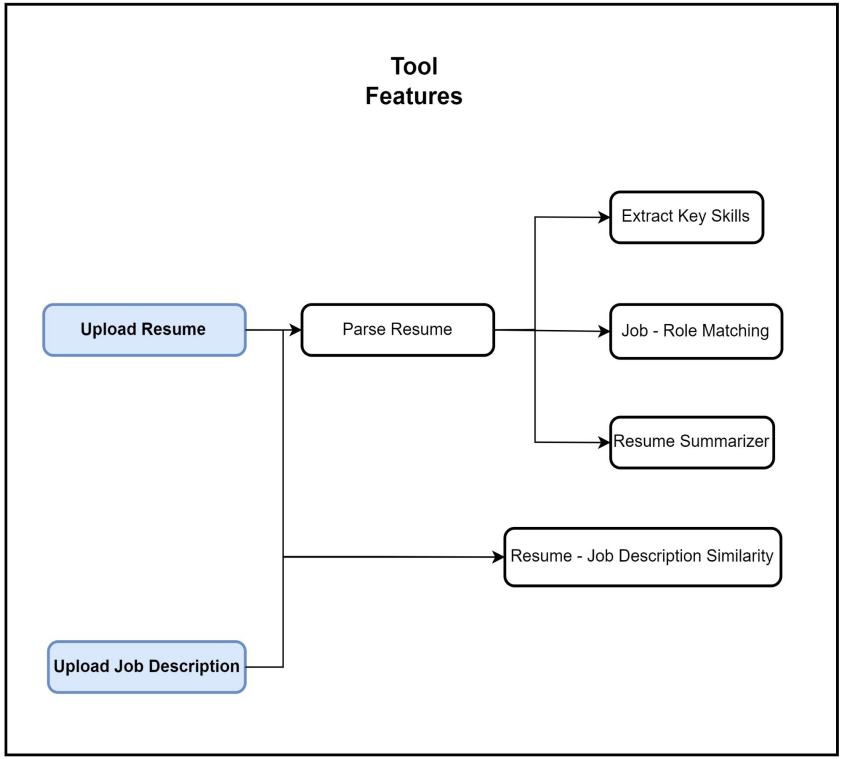
Data Description

Multi-labeled dataset of resumes Labeled with occupations

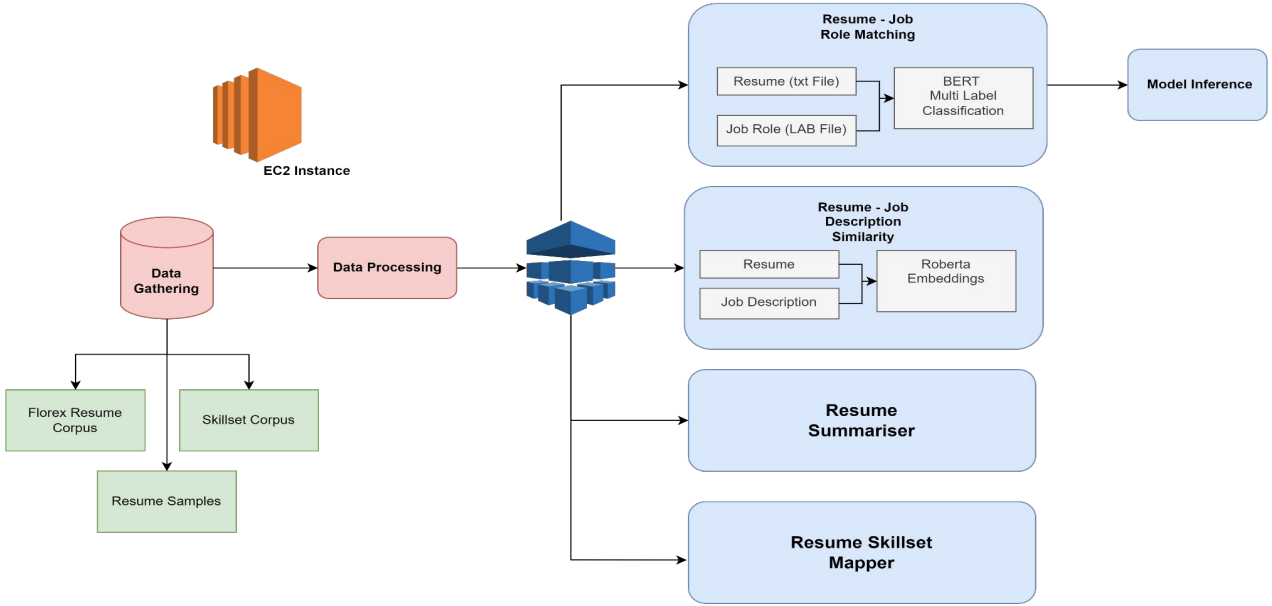
1. Resumes_corpus.zip:
    - Contains individual resumes files in ".txt" format.
    - Corresponding labels are in a ".lab" file.
2. Resumes_sample.zip:
    - Aggregated dataset in a single text file.
    - Structure: Reference ID; Occupations; Text Resume.
    - Occupations separated by ";" in the second field.
3. Normalized_classes: Associations between original occupations and normalized forms.

Features

1. Skillset Extraction
2. Job - Role Matching
3. Resume Summariser
4. Resume - Job Description Similarity



Experimental Setup



Resume - Job Role Matching

- Objective: The code aims to map resumes to job labels using a Machine Learning Model.
- Components: Utilizes the Traditional ML and Transformers library for BERT, PyTorch for model training.
- Data: Custom dataset loaded from a CSV file containing cleaned resumes and corresponding job labels.

Resume - Job Role Matching (continued)

Base Model : Multinomial NB

- Probabilistic algorithm suitable for text classification tasks.

Resume - Job Role Matching (continued)

Data Preprocessing

- Tokenization: Resumes are tokenized using the BERT tokenizer with a specified maximum length.
- Label Encoding: Job labels are processed using MultiLabelBinarizer, converting them into a binary matrix.
- Train-Validation Split: Data is split into training and validation sets (80% training, 20% validation).

Model Architecture

- BERT-based Model: Utilizes the pre-trained BERT model for sequence classification (BertForSequenceClassification).
- Labels: Number of output labels determined by the number of unique job classes in the dataset.
- Optimizer & Loss Function: Adam optimizer with BCEWithLogitsLoss as the loss function.

Results

Model	# Epochs	# Parameters	Precision	Recall	F1
BASE MODEL - MultinomialNB	-	-	54.70%	85.29%	66.65%
BERT Base - Linear Head	10	109.5M	85.15%	78.84%	81.87%
BERT Base - CNN Head	10	109.7M	88.29%	88.29%	85.18%
BERT Base - LSTM Head	10	111.6M	87.09%	82.80%	84.89%

Resume - Job Description Similarity

1. TF-IDF Cosine Similarity:

- Utilizes TF-IDF to represent document content.
- Measures cosine similarity to quantify the similarity between a resume and a job description.

2. BERT and RoBERTa Embeddings:

- Applies pre-trained NLP models (BERT and RoBERTa).
- Tokenizes and computes embeddings for both the resume and job description.
- Uses mean pooling for simplicity in obtaining document-level embeddings.

Resume - Job Description Similarity (continued)

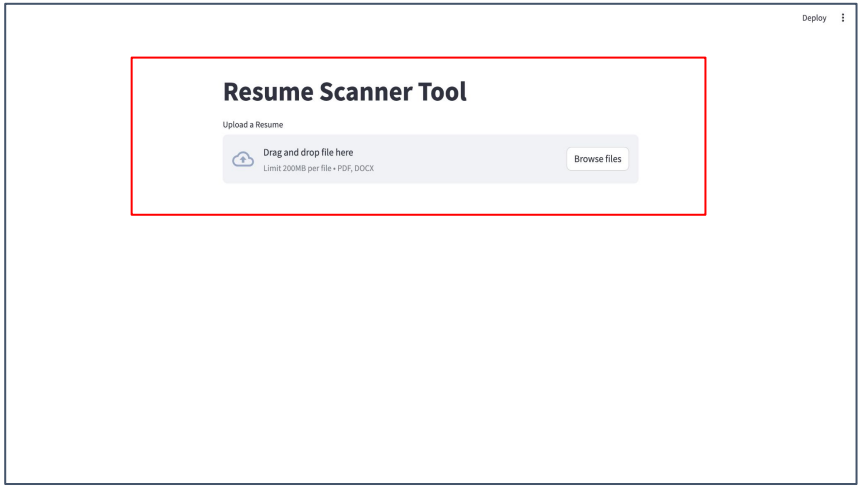
3. Cosine Similarity Using BERT and RoBERTa:

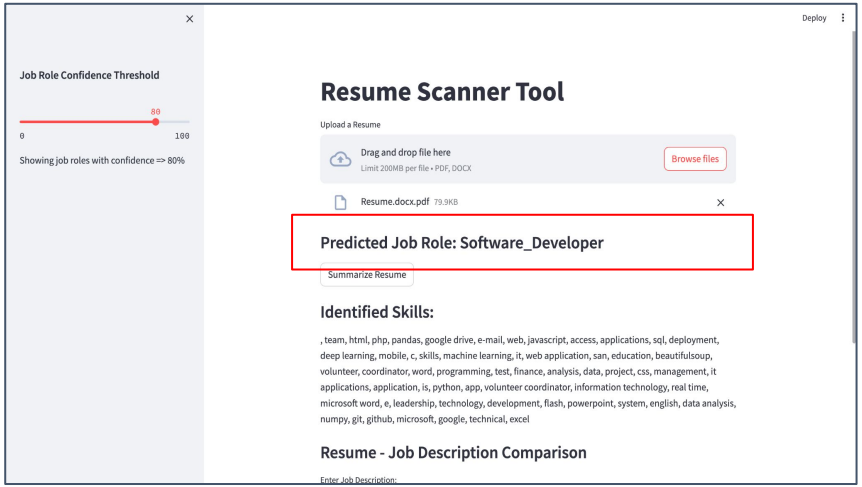
- Calculates cosine similarity between the embeddings of the resume and job description.
- Provides similarity percentages for each model.

Skillset Mapping

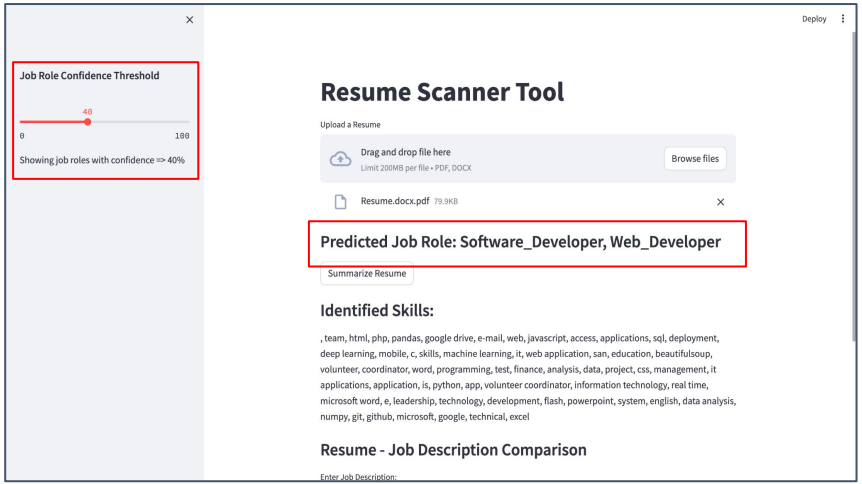
- Step 1: Skills Extraction
  - Extracted skills from the resume_samples dataset.
- Step 2: Master List Creation
  - Processed and compiled a master list of 4000 unique skills.
- Step 3: Skill Matching
  - Matched these skills with the input to determine if the resume contains relevant skills using regular expressions.

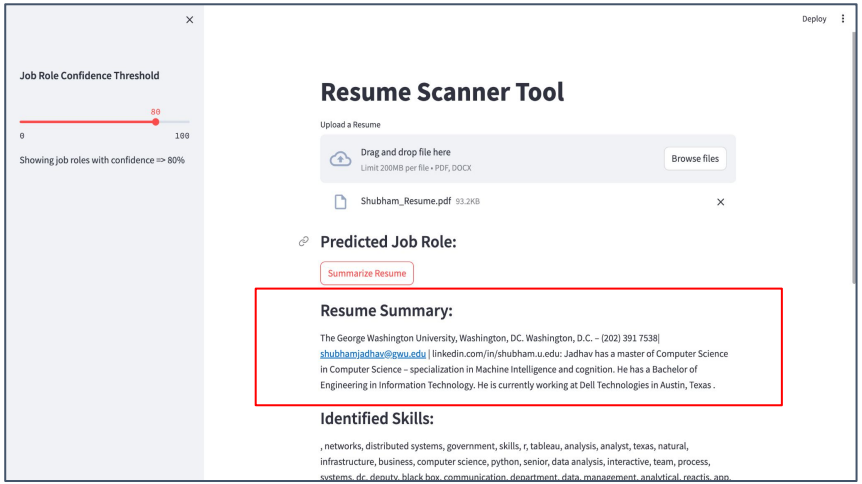
UI and Results



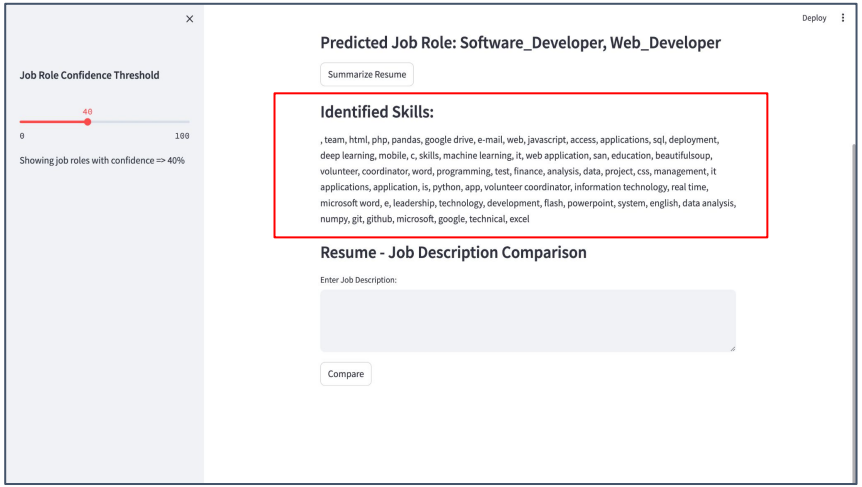


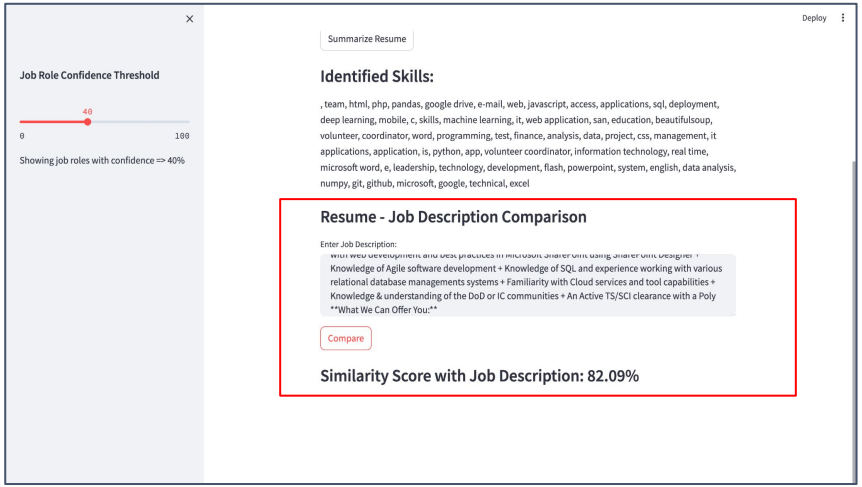
Resume Summarizer





Method: Utilized transformer-based models for automated summarization of resume content.


Deploy
Job Role Confidence Threshold
40
0
100
Showing job roles with confidence => 40%
Predicted Job Role: Software_Developer, Web_Developer
Summarize Resume
Identified Skills:
, team, html, php, pandas, google drive, e-mail, web, javascript, access, applications, sql, deployment, deep learning, mobile, c, skills, machine learning, it, web application, san, education, beautifulsoup, volunteer, coordinator, word, programming, test, finance, analysis, data, project, css, management, it applications, application, is, python, app, volunteer coordinator, information technology, real time, microsoft word, e, leadership, technology, development, flash, powerpoint, system, english, data analysis, numpy, git, github, microsoft, google, technical, excel
Resume - Job Description Comparison
Enter Job Description:
Compare


Deploy
Job Role Confidence Threshold
40
0
100
Showing job roles with confidence => 40%
Summarize Resume
Identified Skills:
, team, html, php, pandas, google drive, e-mail, web, javascript, access, applications, sql, deployment, deep learning, mobile, c, skills, machine learning, it, web application, san, education, beautifulsoup, volunteer, coordinator, word, programming, test, finance, analysis, data, project, css, management, it applications, application, is, python, app, volunteer coordinator, information technology, real time, microsoft word, e, leadership, technology, development, flash, powerpoint, system, english, data analysis, numpy, git, github, microsoft, google, technical, excel
Resume - Job Description Comparison
Enter Job Description:
Knowledge of Agile software development + Knowledge of SQL and experience working with various relational database managements systems + Familiarity with Cloud services and tool capabilities + Knowledge & understanding of the DoD or IC communities + An Active TS/SCI clearance with a Poly
What We Can Offer You:
Compare
Similarity Score with Job Description: 82.09%

Conclusion

- Job Role Matching: BERT Base with CNN classifier head has given the best F1 score of 85.18%.
- Smart Resume Analysis: A Skillset with 400 skills is curated and considered for Skill set Matching.
- Resume - Job Description Similarity : Pretrained Roberta Tokenizer and Model is used to extract embeddings for Resume and Job Description followed by cosine similarity.