Jobs Finder

Group Members

Name	Email
Shoham Weiss	shohamweiss@my.unt.edu
Vrushabh Desai	VrushabhAjaybhaiDesai@my.unt.edu
Harshita Vairagade	Harshitavairagade@my.unt.edu
Sannihitha Kolipaka	SannihithaKolipaka@my.unt.edu

Github

https://github.com/ShohamWeiss/NLP-Job-Finder

Project Description

- Creating several machine learning models to categorize job descriptions and comparison.
- Input is the pure text from the job description.
- Output is a category for job types e.g. Job Title, Experience Required, and Skills Required.

Goals and Objectives

- Motivation

The challenge of matching companies and workers has been long-unsolved, with inefficient roundabout solutions provided by several vendors. Companies either use external vendors to look for appropriate candidates for their openings or publish jobs on job boards, hoping the right candidate will apply. Candidates either publish their resumes and hope they will be contacted, or they are tasked with searching through many job boards to try and find jobs they are qualified for that match their aspirations. All of these processes are currently fully controlled by and dependent on keyword matching. Instead of companies and potential employees just writing what they are looking for in plain English, they both fill out a set of keywords and hope that they will get a high enough match. We suggest applying language models to this problem. The language model would be able to understand plain English job descriptions and match them with plain English applicant information. This project will focus on a subset of this solution, which is categorizing job descriptions using machine learning and comparing several models'

performances, improving on the best model. This will prove the feasibility of the larger vision of having a language model fully understand job descriptions.

- Significance

Job boards currently use keyword matching to match searches and job descriptions. Other machine-learning approaches to this problem have been suggested [1] [2] [3], but they are mostly focused on single categories and do not compare the performance of multiple models. This project will build on previous projects by comparing several machine learning models and improving on the most accurate ones.

- Objective

The objective of the project is to compare several machine learning models trained to classify job descriptions. On top of that, the most accurate model will be improved. The models used in the project will be Naive Bayes, Support Vector Machine, Classic Neural Network, Multilabel Classification with Neural Network, Recurrent Neural Network, Custom Transformer, Custom Transformer on Multiple Tasks, and pre-trained Transformer. The task the models are trained on is taking a text description of a job (taken from a job board) and categorizing it into a set of categories, such as experience level required, job title, skills required, job locations, and more.

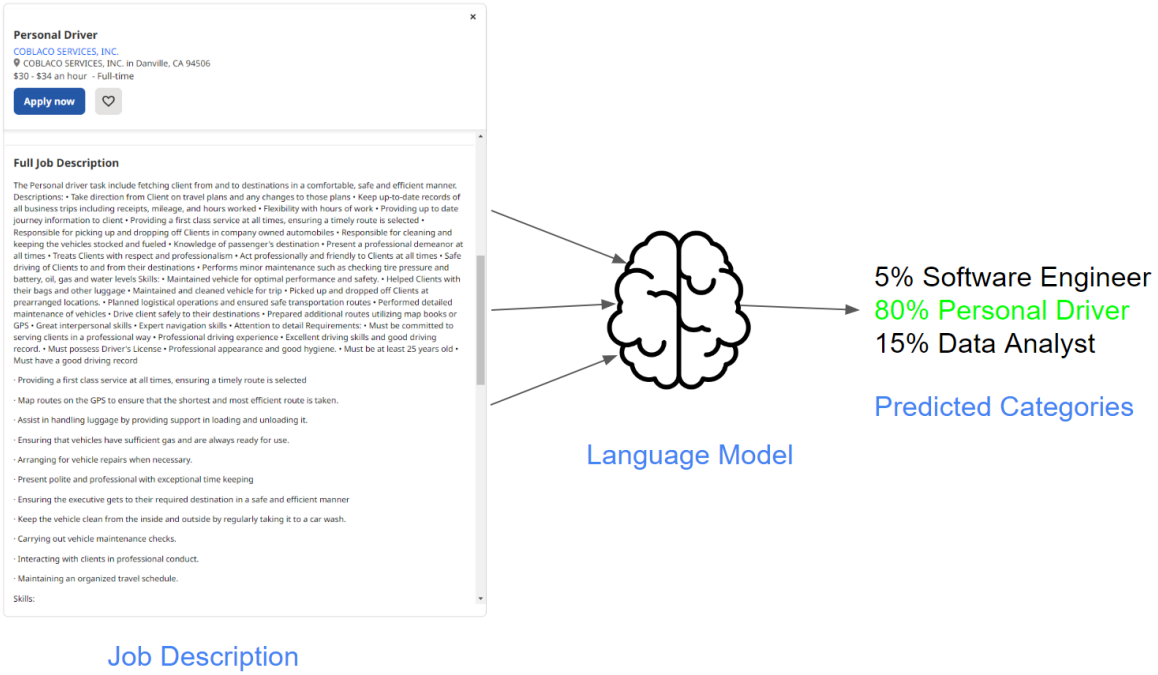

Figure 1: A simplified diagram showing the flow of data in our project. On the left a job description is provided as input to a language model in the middle which outputs a predicted category on the right

- Features
    - Scraping job boards for job descriptions and logically labeling them (multilabel, job title, years of work experience required, level of education required, skills list, location)
    - Run Exploratory Data Analysis on the scraped data. (Show num of desc. in each category, show the most common non-stop words in each category, word clouds,
    - Training several machine learning models on the job descriptions and comparing results. Each model type will have 3 different variations one for each classification type
    - Improving on our best model

References

1. *Job search | indeed*. (n.d.). Retrieved March 8, 2023, from https://www.indeed.com/
2. Sanabdriss. (2020, November 24). *NLP: Extract Skills from job descriptions*. Kaggle. Retrieved March 8, 2023, from https://www.kaggle.com/code/sanabdriss/nlp-extract-skills-from-job-descriptions
3. Jain, A. (2018, October 14). *Classifying job posts via NLP*. Medium. Retrieved March 8, 2023, from https://medium.com/data-science-101/classifying-job-posts-via-nlp-3b2b49a33247