Project Proposal

Group Members:

Jeff Mathew Sam, Christopher White, Josh Li, Lansing Wilson

Dataset:

https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024/discussion/477663

Group Rational:

As a group of graduate students, we decided that one of the most meaningful models that we could produce was one to help those in need of a job. The premise of our model is to assist individuals who are attempting to enter the job market post-graduation. The focus area within our model is to identify keywords from a LinkedIn dataset that we recovered from Kaggle.

Our group would like to explore the possibility of individuals searching for a job, sector, or geographic location of their interest and using text recognition to pull out trending keywords from the job_skills and job_summary of jobs posted on LinkedIn. The purpose of these keywords is to assist in optimizing an individual's resume. Our goal is to encourage individuals to aim for a job one level above the one they are applying for. Therefore, our dataset will show Data Science positions for Senior and Associate level positions.

A special consideration that we have taken into our research is the actual (functional) link within LinkedIn for individuals to conduct their exploratory research into the job or sector of their choosing. If this model is successful, and with additional datasets, we can expand to a much larger array of postgraduate jobs, locations, and nuanced attributes. An additional consideration we can make for future refinement would be to add a feedback loop. This would help us quantify the success of our code and how to alter the code in any meaningful way.

Key Features of Dataset:

- Analyze the most in-demand job titles for data science or data analyst professionals based on job postings on LinkedIn in 2024.
- Identify the top companies hiring for data science positions from 1.2 million job listings.
- Determine the most common job locations for data science roles.
- Explore the relationship between job level and required skills.

Limitations of Dataset:

- Data only depicts a “job level” of two (Senior and Associate)
- Data is limited to the jobs posted between 11 Jan - 19 Jan 2024
- Data shows 4 out of 17 columns after merging would not be considered useful data

Future research:

- Find and gather salary data if available for the job postings to find expected remunerations job seekers can expect in specific fields in the current market conditions.