

Spring 2024 Big Data: CSGY 6513-D

Name: Jiaxin Dong, Chenchen Guo, Mohammed Zakriah Ibrahim

NetID: jd5226, cg4421, mi2471

Big Data Project Proposal

Abstract

In the rapidly evolving job market of 2024, understanding the dynamics of job availability, industry demands, and skill requirements is crucial for both job seekers and employers. This project proposes to utilize a comprehensive dataset containing 1.3 million job listings scraped from LinkedIn, augmented with detailed job skills information, to gain insights into the current job market trends, identify skill gaps, and develop a job recommendation system. The dataset, a rich source of information on job titles, industries, companies, and required skills, offers an unprecedented opportunity to analyze and address the needs of the modern workforce.

Statement

Our project will be centered around several key objectives, utilizing the dataset to conduct a thorough job market analysis, identifying trends in job titles and industries across different geographies. Map out the skills landscape, pinpointing the skills most frequently listed in job postings across various job categories. Develop a sophisticated job recommendation system that matches job seekers with suitable positions based on their profiles, experiences, and skill sets. Explore patterns and correlations between job types, levels, and required skills across different industries, providing insights into the qualifications needed for career advancement in various fields. Inform educational and training programs by identifying current skill gaps. Investigate the relationship between job titles and required skills, offering a nuanced understanding of how specific roles are evolving in terms of skill requirements.

Objective

Exploratory Data Analysis (EDA) on Job Market Data: Perform a comprehensive exploratory data analysis on job market datasets to uncover underlying patterns, detect anomalies, and gain insights into the job market dynamics. This topic could cover visualization of data distributions, identification of key variables influencing job market trends, and preliminary assessments of data quality and structure.

Skills Mapping: Leverage the skills data within the dataset to map out the most in-demand skills across different sectors. This analysis will identify core competencies sought after by employers, facilitating a better alignment between job seekers' skill sets and market needs.

Job Recommendation System Development: Develop a sophisticated job recommendation system with scoring search algorithms that matches job seekers with potential job listings based

on their profiles, previous experience, and skill sets. This system will aim to streamline the job search process and increase the chances of successful employment.

Industry and Job Title Trends: Identify emerging trends in job titles and industries, spotlighting growth sectors and roles that are becoming more prevalent. This will help job seekers and educational institutions tailor their focus towards areas of future demand.

Temporal Trends Analysis: Evaluate the frequency of job listings over different time frames (daily, weekly, monthly, annually) to identify trends and patterns. Determine peak seasons for job listings in various industries to help job seekers optimize their search strategies and companies to plan their hiring cycles.

Job Title and Skills Relationship Exploration: Investigate the relationship between job titles and required skills to uncover the specific competencies that are critical for success in various roles. This will help job seekers focus their skill development efforts more effectively.

Geographical Data Analysis in Job Markets: Delve into the analysis of geographical data related to job markets to understand regional employment trends, economic activity, and demographic impacts on job availability and industry growth. This would include exploring data on job density by region, popular job sectors per geographical area, and migration trends of skilled labor.

Correlation Analysis Between Job Skills and Job Titles: Investigate the correlations between job skills and job titles to determine how specific skills are aligned with certain job roles. This analysis would help in understanding the direct relationships and dependencies between the skills demanded by employers and the titles of the jobs posted, which could be pivotal for both job seekers and recruiters in terms of training and hiring respectively.

Methodology & Technology

Data Cleaning and Preprocessing: Clean the dataset to ensure accuracy and usability for analysis. This includes handling missing data, normalizing job titles, and categorizing skills.

Statistical Analysis and Search Algorithms: Utilize statistical methods and features, develop scoring search algorithms to conduct comprehensive analyses and develop the job recommendation system.

Visualization and Deployment: Create intuitive visualizations to present findings, trends, and recommendations clearly to a broad audience, including job seekers, employers, and policymakers. A web based GUI client side app for easier use.

- Python 3.10+
- Pandas
- PySpark
- Jupyter Notebook

- Matplotlib
- Seaborn
- Streamlit

Data source:

1.3M LinkedIn Jobs & skills

Dataset size: 2GB

Number of Record: 1,296k

Link:

<https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024/data>