Job Market Analysis

Course Name:
SP25_CSGY-6513-D_Big Data

Semester:
Spring 2025

Members:
Ilka Jean ifj2007
Neha Nainan nan6504
Shruti Pangare stp8232

Project Abstract
The platform will analyze historical and real-time job market data across multiple dimensions - geographic, temporal, industry-specific, and skill-based to generate high-precision forecasts of emerging opportunities, fading roles, and evolving skill requirements. Our approach combines time-series analysis, and predictive analytics to identify patterns invisible to conventional labor market analysis methods. By democratizing access to sophisticated job market intelligence, we aim to reduce skill mismatches, enhance labor market efficiency, and promote more equitable economic opportunity in an increasingly volatile employment landscape. The platform will empower job seekers to make data-informed career decisions, help employers develop forward-looking talent strategies, and provide policymakers with evidential foundations for workforce development initiatives.

Problem Statement & Objectives
Problem Statement:
The job market, including engineering, finance, and corporate sectors, has been affected by economic shifts, automation, and changing workplace dynamics. This project will utilize big data analytics to examine hiring trends, salary fluctuations, and emerging job roles across multiple industries. By analyzing job postings, labor market reports, and economic data, we aim to provide a predictive model for job market stability and growth.

Objectives

Our project aims to develop a comprehensive job market analysis platform offering accurate and accessible employment insights to stakeholders. The core objectives of this endeavor are as follows:

1. **Data Processing and Analysis**:
    - Collect, clean, and preprocess historical job posting data from multiple sources using PySpark.

    - Extract meaningful features from job descriptions, including skills, qualifications, industry sectors, and geographic locations.
2. **Predictive Modeling for Job Market Trends:**
    - Develop machine learning models to forecast job market trends based on historical job posting data.
    - Evaluate models using performance metrics and iterate to improve prediction accuracy.
3. **Unified Dashboard Development for Comprehensive Analysis**:
    - Create an intuitive and user-friendly job market analysis dashboard that delivers real-time and future employment forecasts from multiple models.
    - Allow users to view and compare predictions across different dimensions (skills, industries, locations) on a single platform, facilitating informed decision-making for career planning and hiring strategies.
4. **Skill Gap Analysis and Career Recommendations**:
    - Identify skill gaps between current workforce capabilities and emerging market demands.
    - Provide personalized career transition recommendations based on transferable skills and projected market opportunities.

Data Sources:

https://www.kaggle.com/datasets/ravindrasinghrana/job-description-dataset
Dataset File Size: 1.74 GB
No of records: 1615940 records

https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024
Dataset File Size: 6.19 GB
No of records: 1.3 million records

Proposed Technologies and Programming Languages:

Python, Tableau, Apache Spark, Google Cloud
Train ML models for trend prediction using Spark MLlib
Develop NLP pipelines for skill extraction using NLTK/spaCy
Implement time series forecasting with TensorFlow