MATH 485 – FINAL PROJECT PROPOSAL

Title of the Project

Job Market Analysis Using Clustering, Salary Prediction, and Skill Network Analysis.

Team Members

1. Pushpak Sunil Rane
2. Amol Bhalerao
3. Vatsal Rami

Brief Description of the Topic

In today's fast-changing job market, understanding hiring trends, required skills, and salary expectations is crucial for both job seekers and employers. This project aims to analyze job postings to extract meaningful insights, such as job demand across industries, salary predictions based on job descriptions, and skill co-occurrence patterns. We will leverage clustering techniques to group job roles, regression models to predict salaries, and network graphs to visualize skill relationships. The dataset will be collected from free job listing APIs. The insights generated from this analysis will be valuable for students, professionals, and recruiters in making informed career decisions.

Expected Outcome of the Project

- Categorization of job postings into clusters based on job roles and skills.
- A machine learning model to predict salaries based on job title, skills, and location.
- A network graph showing the relationship between skills frequently appearing together in job postings.
- An interactive dashboard to present findings effectively.

Initial Plan

1. **Research (Week 1-2)**
    a. Identify suitable job market APIs for data collection.
    b. Review clustering, regression, and network analysis methods.
    c. Gather sample datasets for testing.

2. **Design and Prototyping (Week 3)**
    a. Define the data schema and preprocessing steps.
    b. Set up clustering techniques for job categorization.
    c. Build and test initial salary prediction models.
    d. Design the structure for the skill network graph.

3. **Data Collection, Analysis, and Reporting (Week 4-5)**
    a. Collect and clean job data from APIs.
    b. Train and fine-tune the clustering and regression models.
    c. Create network graphs for skill visualization.
    d. Prepare final report and project demonstration.

Task Decomposition Per Team Member

Member 1 –

- Data Collection and Preprocessing, Salary Prediction Model Development
- Identify and integrate APIs for job data collection.
- Write scripts to fetch and clean data (handling missing values, duplicates).
- Feature engineering for salary prediction (location, experience, job type).
- Train and evaluate regression models for salary estimation.
- Document findings and model performance.

Member 2 –

- Job Clustering and Skill Network Analysis
- Implement clustering algorithms (K-means, Hierarchical Clustering) to group jobs.
- Analyze job descriptions to identify key required skills.
- Build a skill co-occurrence network graph.
- Optimize cluster performance using feature selection techniques.
- Document clustering insights and trends in job postings.

Member 3 –

- Dashboard Development and Report Preparation
- Develop an interactive dashboard for data visualization.
- Integrate job clusters, salary insights, and skill networks into the dashboard.
- Implement data visualization components:
- Job Market Overview: Bar charts and heatmaps for job demand trends.
- Job Clustering Insights: Scatter plots and cluster analysis visualization.
- Compile the final report summarizing methodology, results, and insights.
- Prepare slides and demonstration materials for the presentation.