**Project Outline: Finance Job Scraping & Forecasting System**

### **1. Project Overview**

This project aims to develop an automated system that scrapes finance-related job postings, analyzes job descriptions against a general-purpose finance resume, classifies job postings by sector, and forecasts future job opportunities. The system will leverage NLP, SQL databases, and Streamlit for visualization.

- Need to purchase Ultra JSEARCH API for 50,000 per month

### **2. Tech Stack & Tools**

* **Job Scraping:** JSearch API
* **Data Storage:** SQL (SQLite)
* **Text Analysis:** NLP (Regex, TF-TDM Model)
* **Sector Classification:** YFinance API
* **Forecasting:** ARIMA (AutoRegressive Integrated Moving Average)
* **Visualization:** Streamlit
* **Development Environment:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)

### **3. Project Workflow**

#### **Step 1: Job Scraping with JSearch API**

* Pull (50000) job postings related to finance using JSearch API.
* Extract relevant fields: job title, company, location, description, and ticker (if available).
* Store data in SQL database.

#### **Step 2: Resume Parsing & Keyword Matching**

* Load a general-purpose finance resume (text or PDF).
* Use NLP and Regex to extract key descriptive words from the job descriptions.
* Implement a TF-TDM model to analyze frequency of key skills and terms.
* Rank job postings based on similarity to the resume.

#### **Step 3: Sector Classification Using YFinance**

* Extract tickers from job postings.
* Use YFinance API to fetch sector and industry classification.
* Categorize job postings into finance sub-sectors (investment banking, asset management, fintech, etc.).

#### **Step 4: Historical Job Posting Tracking**

* Maintain a SQL database of past job postings.
* Track trends in job openings over time based on keywords and classifications. (Quarterly)
* Identify hiring patterns and job market shifts for Job Titles.

#### **Step 5: Forecasting Future Job Postings**

* Use historical job posting data to train an ARIMA time-series forecasting model.
* Forecast job postings by sector and keyword trends.
* Generate predictive insights on hiring demand for Job Titles.

#### **Step 6: Visualization with Streamlit**

* Develop visualization displaying:
  + Current job postings and their sector classification.
  + Resume-job similarity scores.
  + Historical job trends and sector-wise hiring patterns.
  + Forecasted job opportunities using ARIMA.

### **4. Next Steps & Timeline**

|  |  |
| --- | --- |
| **Task** | **Deadline** |
| Set up JSearch API integration | Week 1 |
| Develop SQL database schema | Week 2 |
| Implement NLP text extraction | Week 3 |
| Integrate YFinance for sector classification | Week 3 |
| Train ARIMA model for forecasting | Week 4 |
| Build Streamlit dashboard | Week 5 |
| Final Testing & Optimization | Week 6 |

### **5. Expected Deliverables**

* A functional job scraping pipeline.
* SQL database storing job postings and historical trends.
* NLP-based resume-job matching system.
* Sector classification based on company tickers.
* Forecasting model for job trends.
* Interactive Streamlit visualtizations.