**Requirements**

**Part 1**

1. Create a database of employer websites which include employer name, employer description, employer location, employer country, Industry, career page link etc. using scrapping which can be used further to increase employer database.

e.g. CV Library, Indeed etc. can be used to create employer website and scrap employer website to find out career page link, location etc.

1. An Api need to create which will be called from other server with parameters like Job title, industry, salary, job type, job location etc. to find out matching jobs from employer website career pages.
   1. Create a web scrapping bot
      1. Scrap employer career page as location mentioned in employer database considering all type of page development like listing, pagination, ajax based, search method, page inheritance etc.
   2. Process Scrap pages using ML Model
      1. There can be different technologies/hierarchy used on career page (Consider all type of premutation and combination to scrap web pages)
      2. Career page can have multiple pages, search fields, ajax etc. to go to next page (Please consider different way to scrap career pages)
      3. Remove un-necessary data from scrap pages and find out multiple jobs if there from single page and find out “apply now” link with each job
   3. Match jobs with provided criteria’s in api using AI
      1. Match jobs with provide criteria’s/fields using AI with accuracy
      2. Return higher accuracy matched jobs in descending order
2. Cache Results and employer website career pages for a day to process faster the results
3. ML Model accuracy should be > 90% and speed should be enough for real time processing for large number of employer websites.
4. Deploy on AWS EC2 using Load Balancer.

**Part 2**

Match the candidates resume (doc file or in database) with provided criteria’s and return result with accuracy using AI and ML Model

Skills Required:

* Scrapping Library like Scrappy/Panda
* Mathematical and statistical knowledge (e.g. linear algebra, probability, statistics)
* Programming knowledge (preferably in Python)
* Knowledge of ML algorithms and libraries (e.g. Tensorflow, PyTorch, scikit-learn)
* Data cleaning, pre-processing, and feature engineering
* Experience with data visualization and analysis
* Understanding of overfitting, regularization, and evaluation metrics
* Ability to perform error analysis and model selection
* Understanding of the business problem and the ability to translate it into an ML problem.

**Project Modules:**

1. Web Scraping Bot

2. Data classification using ML

**Web Scraping Bot:**

Need improve the existing Web Scraping bot which will:

1. Scrape the HTML source of multiple jobs

2. Call the ML classifier

3. Fetch the response and store it in a database/CSV files

**Data classification using ML**

Develop a Python Machine Learning based Classifier that will:

1. Prepare the dataset

2. Analyze and clean it

3. Fetch the dataset prepared

4. Train the classifier based on that dataset

5. Predict/classify the parameters from the provided HTML source

6. Store the output

7. Calculate the precision

8. Be integrated with the Web Scraping Bot

**Steps:**

1. dataset that needs to be labeled to be used by the ML classifier.

2. After the dataset, a classifier will be trained (a good GPU is required for this usually)

3. Once the classifier is trained, develop it in such a way that it can act as an API so it can be integrated with the Web scraping bot.

4. Finally, integrate everything on an AWS EC2 instance to run 24/7

**Tech Stack:**

1. Python

2. NLP

3. Tensorflow

4. BeautifulSoup/Selenium

5. Requests

6. Numpy

7. CSV

8. Concurrent futures (for multithreading), etc