Assignment

1. A developer is assigned a task to scrape 1 lakh website pages from a directory site, while scrapping he is facing such captcha, which are placed to stop people from scrapping As a project Coordinator suggest ways to solve this problem.

Sol:-

* Using Selenium to automate instructions with the website. These applications have the capability to replicate human interactions and potentially solve CAPTCHAs in a manner akin to human interaction.
* Developing a machine learning model for image recognition to solve CAPTCHAs by using a dataset of CAPTCHA examples to train the model. It might be more complex to implement.
* Checking if the website provides an official API that van be used instead of scraping data directly.

2. Our client has around 10k LinkedIn people profiles, he wants to know the estimated income range of these profiles. Suggest ways on how to do this?

Sol:-

* Analysing the job titles and industries of the profiles. Websites like Glassdoor can provide average salary data by job title and location.
* Correlate education levels, years of experience and professional qualifications. These could provide indications regarding likely income brackets in accordance with industry standards.
* Machine learning models, if trained on relevant datasets, could predict income ranges based on profile information.

3. We have a list of 1L company names, need to find LinkedIn company links of these profiles, how to go about this?

Sol:-

* Using web scraping tools to automate web browsing to extract company profiles from LinkedIn.
* Search Engines can be employed to find LinkedIn profiles by querying the company names along with “LinkedIn” to focus results on LinkedIn.
* If available, utilize LinkedIn’s official API to search for company profiles based on company names.

4. How to identify list of companies whose tech stack is built on Python. Give names of 5 companies if possible, by your suggested approach.

Sol:-

* Searching for job openings or careers section on a company’s website as many companies mention the technologies they use.
* Explore GitHub repositories. Companies often make their code open source which can be a good indicator of their tech stack.
* Companies might announce their technology choices in press release or public statements.

Some Companies that use Python extensively are –

* Google
* Instagram
* Spotify
* Facebook
* Netflix

5. Need to find an API, through which we can send LinkedIn messages to other LinkedIn users.

Sol: - LinkedIn’s available API does not support sending direct messages to other LinkedIn users, as it imposes limitations on accessing private messaging functionalities.