Task 2 Project Template

<https://isthisarealjob.com/>

Data privacy and piracy remains a big issue in today’s world, lately tech giants like Facebook and google have been faced with backlash and fines from government institutions and their users, these has prompted a reordering of data collection and handling.

From the above, there is a limitation to how we can go the route of sourcing for users data from third parties as such, the greatest challenge lies in the data collection and handling, recognizing this, below are my propose solution to improve the site.

**Site Develop and Remodeling:** There should be a change in the business model of the site, it as the current model does not have enough attributes to attract users, many users will not have the time or the reason to confirm the authenticity of jobs especially when they have been on the hunt for a long time. Job authentications should be an added feature to the site. In view of this, the site should be remodeled and developed to be a job posting site and employment test site for users, where user can take sample pre-employment tests and interview questions. As these models will attract users who in turn comes with data in form of user’s details login inputs, employers detail inputs amongst others. This will also reduce the manual update of the site database as inputted data will be automatically saved to the database.

I will be working with dedicated UI designers, Back and front-end experts that will use web development tools such as HTML, CSS, JavaScript, php to remodel the site and improve the site’s functionality.

**External data Collection and storage:**  In addition to the data on the database, I will be carrying out sentiment analyses, web data extraction on social media platforms and some trusted. Job platforms using APIs of these social media, beautiful soup library and the spacy library. Data will be extracted base on the data features we have on the database, emphasis will be on keywords such as company name, job title and location.

External extracted data will be merged with our own data in our database

**Data analysis and Job validation:** During the data analysis, each company will be identified with their unique identification/registration number found on the Nigerian Corporate Affairs Commission and tax ID. Job from companies without any unique registration number or tax ID will be tagged as fake while job postings from companies with tax ID and registration number be validated on the sentiment analysis done on platforms such as twitter and nairaland done as at the last three years such jobs postings where made. For a job title an average sentiment of less than 0.2 will be classified fake. 0 to represent fake, 1 to represent real.

Data analysis, grouping and mungling will be done with pandas, numpy and matplolip

**Machine learning algorithm connected to the Database:** A Random forest or K-nearest neighbor classifier algorithm will be trained base on the data collected and analyzed above. To identify the pattern of recognizing fake and real jobs, after which it will be deployed on the site as well as connects with the database.