**TASK:**   
<https://isthisarealjob.com/> is a website that tells a user if a job is real or not. At present, the database is updated manually. Now there limited scope of jobs that will be covered by us. We want to improve this website with machine learning. How can we do this?  
We need to be able to improve the whole process from data collection to prediction of a real job using machine learning. So how will you as an ML Engineer solve this problem for us.

**Proposed Solution:**1. Find out the real job posts as well as fake job posts

a. Run a scraper on known online job ads/services e.g Nairaland, jobzilla.

b. Use BeautifulSoup or Urllib

c. Collect text of fake job posts these and update a database storing the agency names addresses phone numbers and email.

This script would run periodically to commit updated fake job posts to the database

d. Collect text of real job posts from Jobberman, Nairaland

This script would run periodically to commit updated real job posts to the database.

e. Include a field set to 0 for fake and 1 for real job posts

**2. Perform EDA**

Get the most frequently used words:

- Do a CountVectorizer(ngram\_range=(1,1))

- Repeat CountVectorizer for bigrams.

- Make a stop\_words list to be used in modelling our data

3. Use both vectorizers and classification models to find the best parameters that give the highest accuracy. CountVectorizer, Tfidvectorizer, LR and MNB

Now our model has learnt how to classify a job as real or fake.

4. Take the model and deploy.