

Education through collaborating on real world problems



<https://omdena.com/chapters/>

The World's Platform for Impact with AI

Community ● Collaboration ● Impact



Solving community problems using Artificial Intelligence



<https://omdena.com/omdena-chapter-page-nigeria/>

The Project

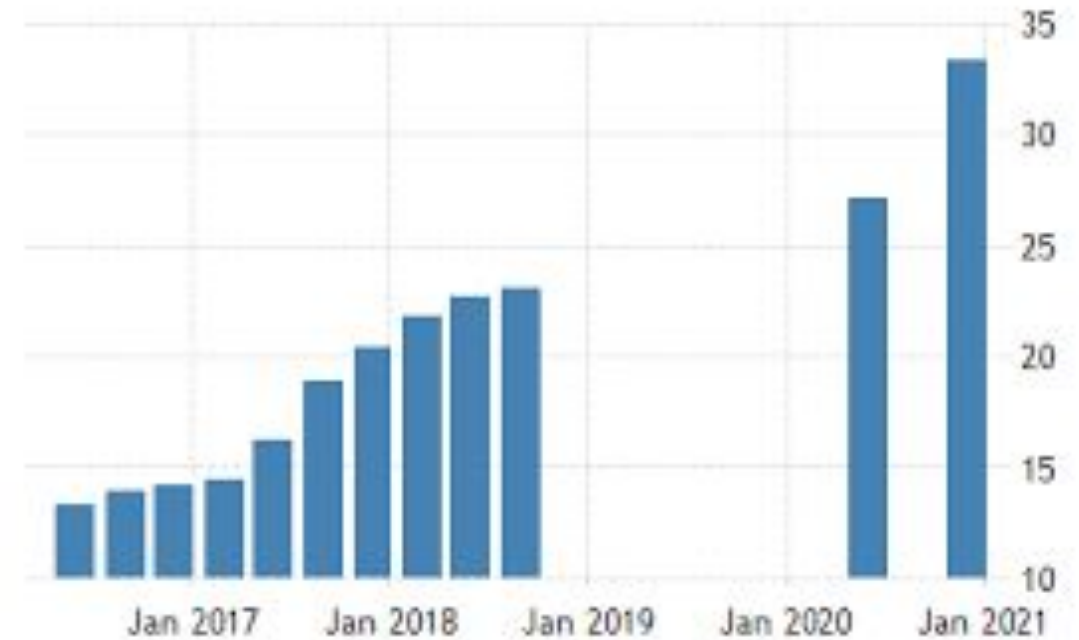


**Solving
unemployment in
Nigeria using Artificial
Intelligence**

Problem Statement



The Unemployment rate in Nigeria increased to 33.30 percent in the fourth quarter of 2020 from 27.10 percent in the second quarter of 2020 while the underemployed rate is 28.7% giving a total percentage rate of 61.9% of Nigerians that are unemployed or underemployed



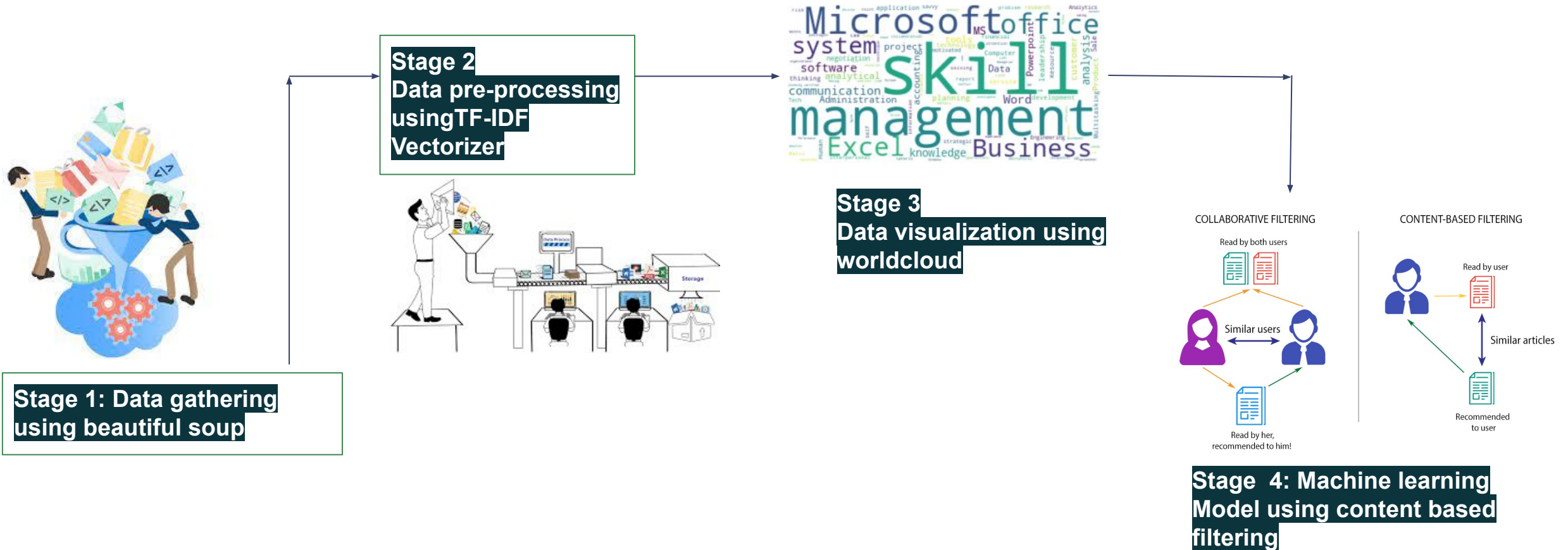
Project Goals



- To create a machine learning model that will merge NYSC graduates with possible job opportunities based on their career path
- To provide employable job skills recommendation to NYSC graduates based on their their career path



Our Approach



Data Gathering



Data source: LinkedIn was the major source where the jobs were scraped.

The **Library used** for the scrapping were:

- Requests
- BeautifulSoup



Data Gathering

A **py script** was created to manage the extraction from the data via web scraping.

Functions methodology used were:

- get_parser function: parse command line arguments
- A function that collects all the jobs url on linkedin
- Function to get the contents from LinkedIn like: title, linkedin_link, Job description, company, location, date_posted.
- Writing the contents to csv



Data Gathering



At the end, comma separate Dataframe we had got had the following features:

- Job title
- Job description
- Date posted
- Link to the Job on LinkedIn
- Company name
- Location
- Skills
- Degree Type
- CVs were also collected from the project members and was used as the graduate profiles



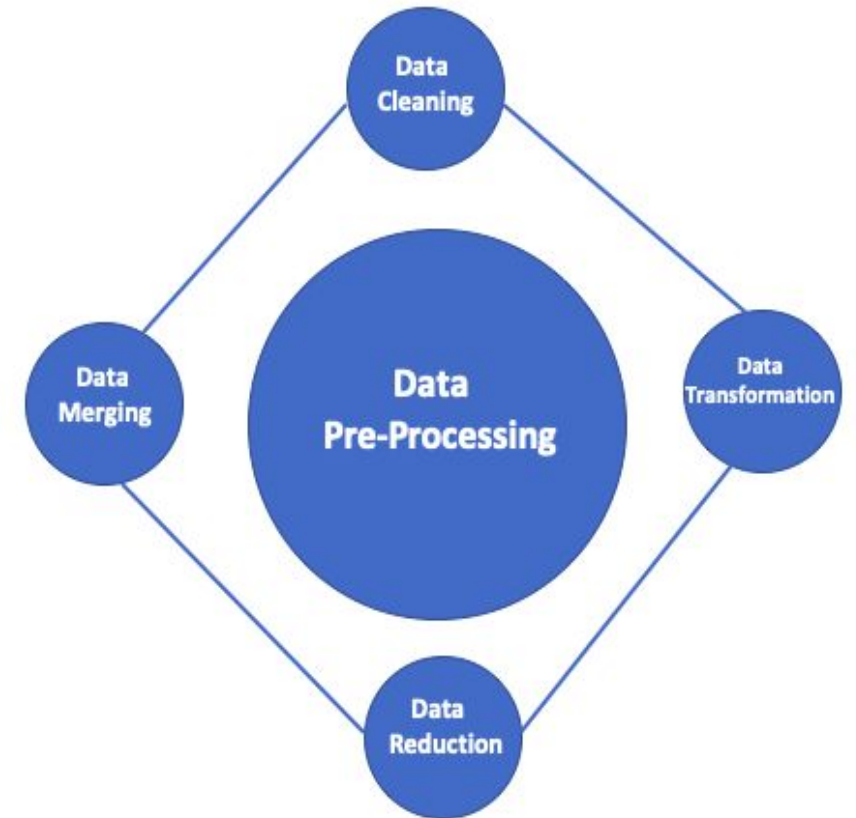
Data Pre-Processing



Data pre-processing can refer to manipulation or dropping of data before it is used to enhance performance. It plays an important role in your model. You need to transform the data in a way the computer would understand and be able to work with it.

Why we need to Pre-process data;

- Accuracy
- Completeness
- Consistency
- Believability
- Interpretability

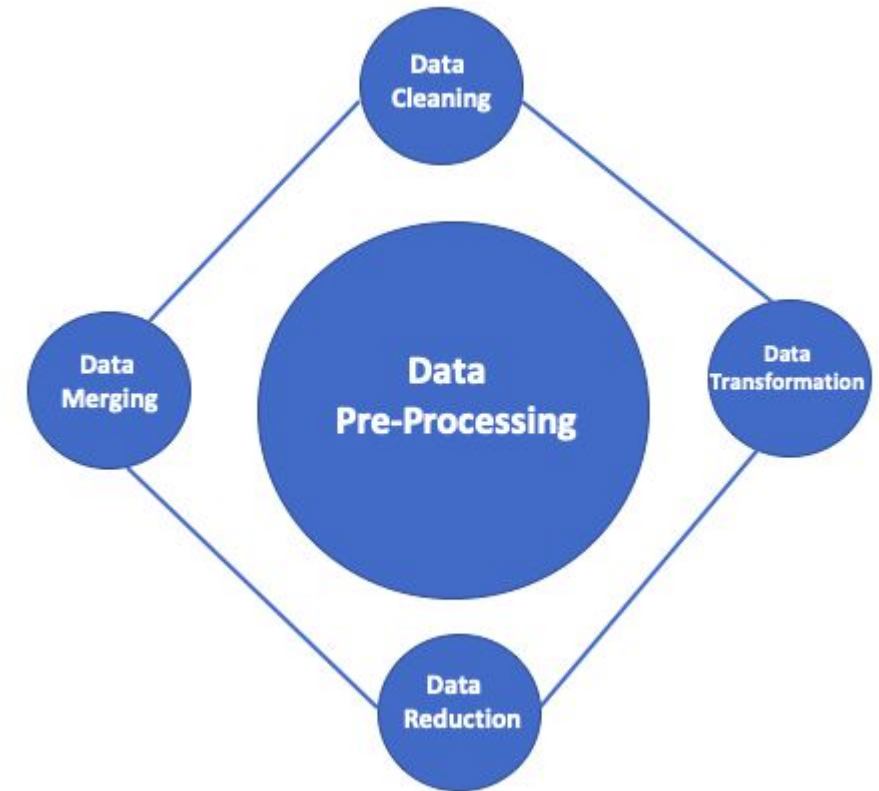


Data Pre-Processing



Data Pre-processing sometimes referred to data preparation involves processes and techniques such as:

- Data Cleaning (Inaccurate/Missing data, Noise and outliers)
- Data Merging (Job dataset and Skill dataset)
- Data Reduction(Duplicate data)
- Data Transformation(Bag of words using the TF-IDF Vectorizer and arriving with an Array)



Data visualization



Steps in building a **WORDCLOUD**

1. Reset the size of the image
2. Remove words not needed in the wordcloud visuals
3. Create a variable name to add the values example all the Job listing
4. Use Join() method to join all the elements together and should be saved in the variable name created above
5. Create an Object of the class Word cloud with the name of your choice
6. Call the generate function () to return the iterable object with a sequence of values
7. Use show () to visualize your data



Data visualization

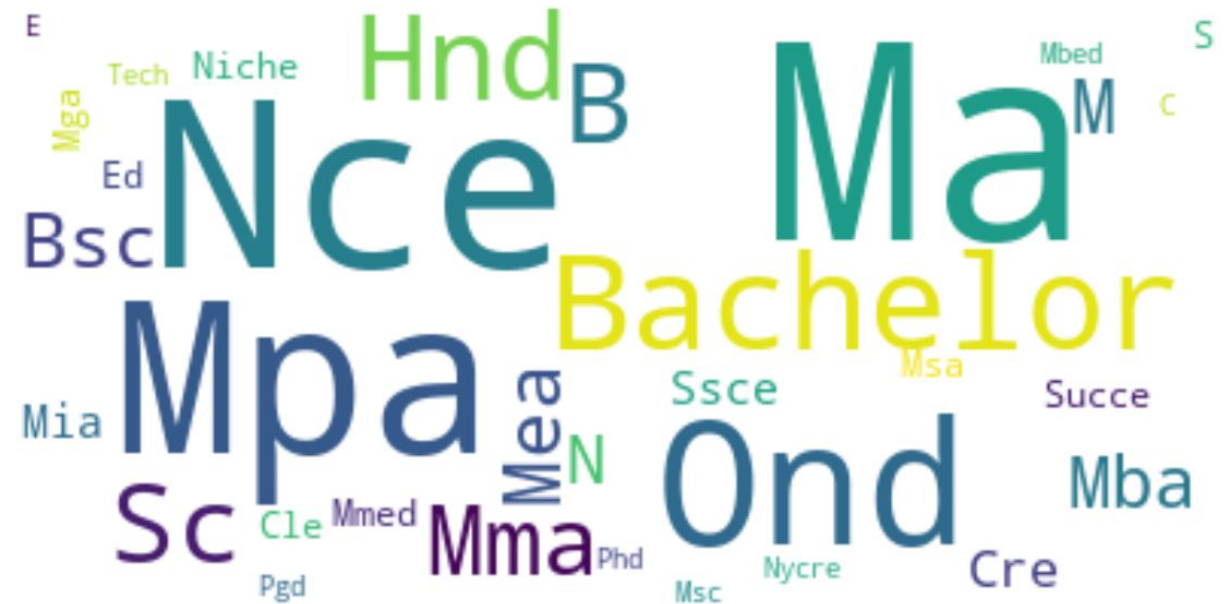


The method and function definition used includes;

join() method: Provides the functionality of joining elements from the iterable separated by a string operator and returns combined string

generator (): is a special type of function which does not return a single value, instead, it returns an iterator object with a sequence of values.

- A **class**: can be defined as a template that describes the behavior the object of its type support.
- An **Object** of the class have the behaviors of their class.
- A **method ()** is an action which an object is able to perform.
- A **function ()** is a set of instruction for a program to perform



Machine Learning Model

There are two major techniques used for building recommendation system.

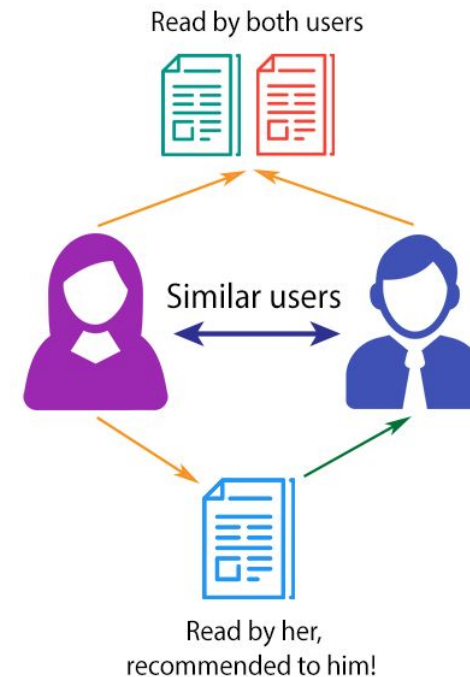
- **Content based technique**
- **Collaborative based filtering technique.**

Content based basically makes recommendation based off on similarities between the content fields. Example, recommending a movie based on it similarities with other movies.

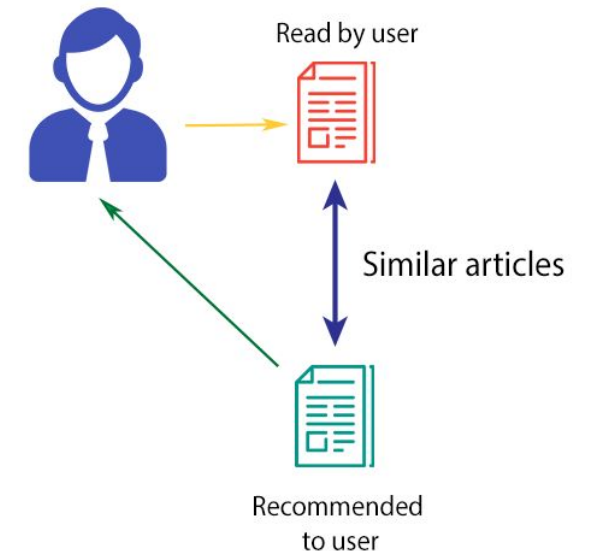
Collaborative based filtering techniques makes use of similarities between users data to make recommendations for new users.

Our approach makes use of the content based filtering as we only had access to public job posting datasets on different sites as mentioned earlier by our data gathering team.

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



Machine Learning Model

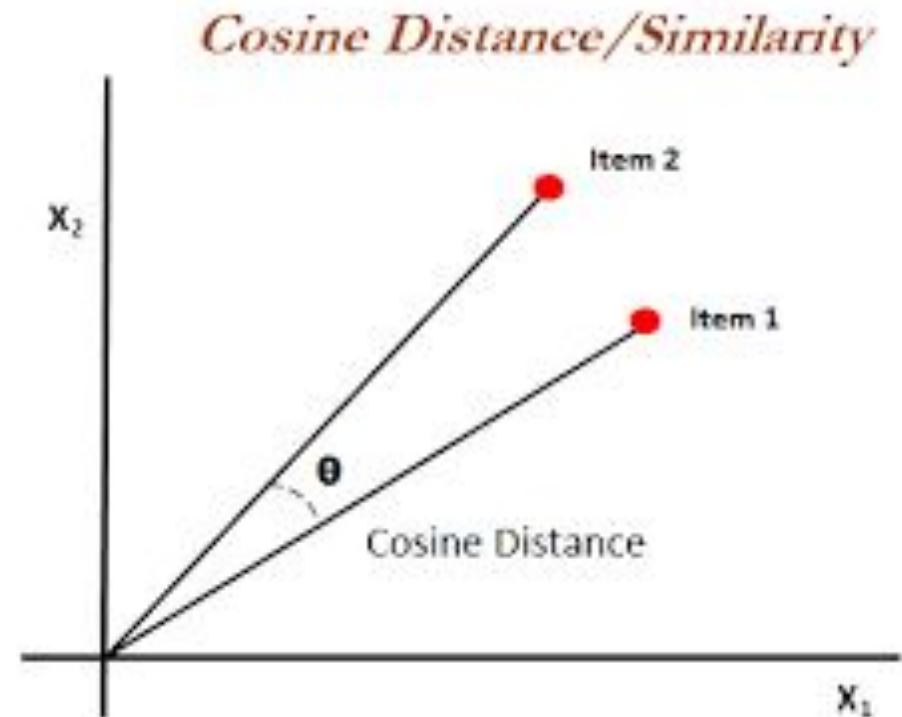
We used two different Algorithms;

- Cosine similarity
- Nearest neighbors.

The **Cosine similarity** makes use of cosine of angles between two vectors to check for most similar items to make it's recommendation. Each record we have on the job posting would be checked with a user's resumé transformed to vectors.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

The values returned by this algorithm ranges from 0 to 1. Values closer to 0 signifies a low similarity between both vectors while values closer to 1 signifies strong similarity between both vectors.



Machine Learning Model

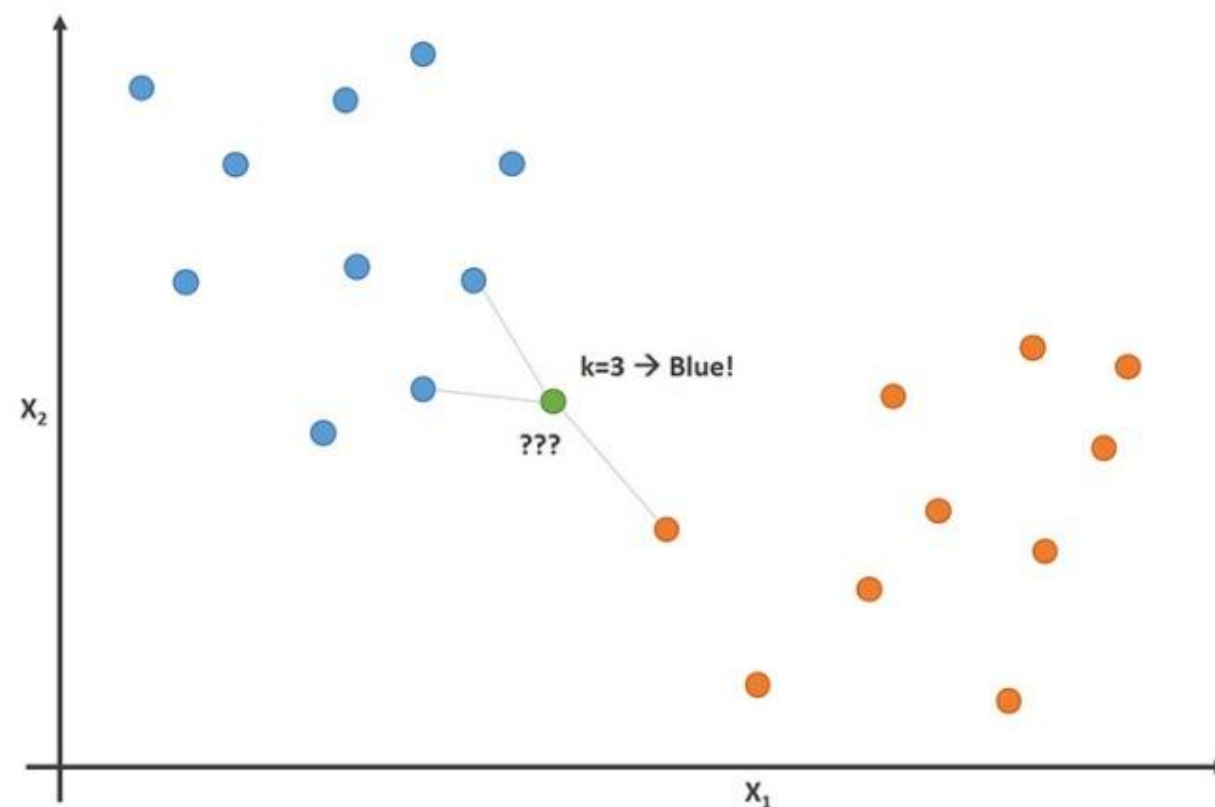
The **Nearest neighbors** implements the content based technique by applying a distance based algorithm, in our case we used the euclidean distance.

$$= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

The **Nearest neighbors** calculate distance between two vectors space. Usually the values ranges from 0 to Infinity. So typically, the lesser the magnitude of the distance between the two vectors, the more likely they are similar and therefore recommended by the algorithm.

But the higher the magnitude the less likely those vectors are similar and would be recommended.



Relevant Links



Omdena Lagos, Unemployment Job scraping URL

<https://www.linkedin.com/jobs/search/?geoid=105365761&location=Nigeria&sortBy=R&start=>

Omdena GitHub Repository

<https://github.com/OmdenaAI/lagos-nigeria-ai-for-unemployment>

Join Omdena Lagos Nigeria Upcoming Project

<https://omdena.com/omdena-chapter-page-nigeria/>

Further Research



For further progress on the work, I'd suggest a neural network for creating word embeddings and expanding the job posting dataset in order to capture a wider range of job postings and requirements for improved performance.



The World's Platform for Impact with AI

Community ● Collaboration ● Impact



Thank You!