TAIYO (PHASE 2)

#1. Scrape data using bs4

#2. Benchmark Accuracy by using :

- Hugging face
- YAKE
- BERT - derived models.
- spacy
- "most population language models" link provided.

} KEYWORD EXTRACTION. OR NAMED ENTITY RECOGNITION.

#3. Provide Visualizations.

THE TASK :-

I (i) Get the data

(ii) use python harvester to scrape or download data.

II USE NATURAL LANGUAGE MODELS WITH TO SATISFY

(1.) To Extract entities :

use Named entity Recognition to identify and extract

- Sector
- Sub-sector
- location

- Government Agency
- Company Name
- Contractors
- Investors

- cost per sq. Km (Unit measurements)

(2.) Similar projects :

using Word2vec / cosine similarity for identifying similar projects within the past 10 years within 500 miles for a given project.

(3) Trends :

To show data visualizations for aggregated time series.

(4) Customize a GPT-3 ChatBot :

A chatbot for QnA for custom dataset.