**Data Analysis To Find Cause And Effect On Real World Events For The Year 2019 – 2020**

Authors – Ankur Choudhury, Ujjal Kumar Maity

Dated 2 January 2023

1. Abstract
   1. Extraction of data from various sources including Twitter, news outlets, etc., using data scrapping methods such as Selenium, BeautifulSoup and APIs.
   2. Process the extracted data. Cleaning and fitting the data into a structured dataset using Pandas, NumPy, etc.
   3. Applying concepts of machine learning and data science to find categories of news or facts from the processed data. Classify the nature of Tweets and the associated user to derive insights about them.
   4. Find a relation between the tweets and real world events. Finding the sequence of occurrence of the events.
   5. Check if the user has any direct or indirect effect on the occurrence of the event(s).
2. Technologies Used
   1. Apache Airflow – Pipelines
3. Resources
   1. Github
      1. HTTPS <https://github.com/ankurchoudhury1/CE001.git>
      2. SSH [git@github.com:ankurchoudhury1/CE001.git](mailto:git@github.com:ankurchoudhury1/CE001.git)
4. Use Case