Big Data Analytics Techniques and Applications
Spring 2023
Term Project

Goal

Prepare a dataset (or multiple datasets) of your interest, set analytics goals, then design and implement an analytics workflow to analyze the selected dataset by using Big Data analytics techniques/tools. The project should be implemented under the following requirements:

1. **This project must be implemented on Hadoop or Spark platform.**
2. **You can use any analytics tools/languages like R, Python, scikit-learn, Spark MLib, etc.**
3. **The selected dataset must carry the characteristics of Big Data (at least one of the Volume and Variety). You should clarify clearly which Big Data characteristics the dataset meets in your presentation/report.**

Dataset

You can use any public dataset (including those crawled from the Web) that meets Big Data characteristics as mentioned above. The following is a good reference source (but not limited):

- Kaggle dataset

***Note: The datasets you select must be publicly available ones. Any proprietary dataset will not be allowed.**

Requirements

1. **Team Size:**
   Default 4 people in each team. (You can seek teammates on E3 Forum <Term Project 隊友招募區>.)
2. **Important Dates:**
   Team registration: **2023/02/27 (M) 10:00:00 – 2023/03/03 (F) 23:59:59**
   Proposal presentation: **2023/03/22 (W) & 2023/03/29 (W) (on-class)**
   - Proposal slides & video upload: **2023/03/20 (M) 23:59:59**

   Final presentation: **2023/05/24 (W) & 2023/05/31 (W) (on-class)**
   - Final presentation slides & video upload: **2023/05/22 (M) 23:59:59**

   Final report uploading due: **2023/06/14 (W) 23:59:59**

*More details and instructions about the proposal and the final presentation will be announced on E3.