Project on Open Data

Data Science

EE5175

About Projects

- 30% of total scores

  – Including 5% for abstract and 5% for presentation

- Open data related
  – With implementation: group of up to 4
- Each team must make a project presentation
  – Dec. 18, 2019 (1st batch)
  – Dec. 25, 2019 (2nd batch)
  – 10 minutes for each team
  – Be precise for the goal and tangible for the progress (at least 50% completed technically)

Open Data Resource

- 政府資料開放平台
- 各直轄市政府的開放資料
- Kaggle
- ESPN
- UN Comtrade
- Others
    - E.g. Stanford Network Analysis Project (SNAP)

What to Expect from Project

- Discover some hidden knowledge from the data you explore (by some mining method(s))

- Clearly describe and demonstrate (by your implementation) what new value is obtained based on the knowledge you discover.

- Abstract due 11:59pm, Nov. 24, 2019 (Sunday)
  - Including: Title, team members, motivation (including related work), method (and tools) to use, expected results
  - One page in Word/PDF

What to Expect from Project (2)

- Final report is due January 16, 2020 (Thursday)
- Within 12 pages (including all)
  - In either Chinese or English

What to Expect from Project (3)

- Content of your final project report

  - The data

    - Your data source and how you process it

  - The tools (open ones)

  - Your methodology

    - Based on at least two of the methods taught in the class: SVD, PCA, Association rule, classification, clustering, etc.

  - Your implementation

  - Your finding

    - Need to be able to explain why the result can be trust

  - The contribution of your work