

Data Science 2021

Term Project Rules

Prof. M.-S. Chen & Prof. M.-L. Lo

Contact TA

❖ Email:

➤ ntuds2021@gmail.com

❖ Ceiba:

➤ forum

➤ Discussion on forum is encouraged.

Outline

- **Rules**
- **Schedule**
- **Grading**
- **Introduction of Datasets**

Rule - team

- 4人一組, 請於**2021/11/24 (Wed.) 12:10前**在以下網址填寫分組資訊
https://docs.google.com/spreadsheets/d/1qVAAtjwcFmrBBHxnw-ZsigVKIJ6x8YwlauVU6sGa_JvY/edit?usp=sharing
 - 需要幫忙分組可寄信到助教信箱
 - 期限後仍未填寫, 助教會幫忙分組
- **組長需填寫 gmail**, 助教會共享 google drive 給組長

Schedule

Date	Todo
11/17 (Wed.) 11:20	final project rules announce
11/24 (Wed.) 12:10	team makeup due
12/8 (Wed.) 23:59	proposal due
12/21 (Tue.) 23:59	presentation slides & short video due
12/22 (Wed.) 9:10	project presentation 1
12/29 (Wed.) 9:10	project presentation 2
1/17 (Mon.) 23:59	report & full video due

- Final exam may be held **physically on 1/5 (Wed.)**, if any.

Grading

- **Project = proposal + presentation + report**
 - Proposal (i.e. abstract) (4%)
 - Oral presentation (including the short video you prepared) (4%)
 - Final report & full video (18%)
- The term project accounts for **26%** of your overall class grade.

Rule - Project topics

- 以下三個資料集擇一
 - Email Spam Classification [[link](#)]
 - Customer Personality Analysis: Analysis of company's ideal customers [[link](#)]
 - LearnPlatform COVID-19 Impact on Digital Learning [[link](#)]
- 根據所選資料集, 題目自訂

Dataset 1: Email Spam Classification [[link](#)]

- **Target:** To predict the 5172 emails of spam or not-spam according to the frequencies of words in the emails
- **Files**
 - **emails.csv:** A CSV file containing spam/not spam information about 5172 emails
 - There are 3002 columns. The first column indicates the Email name.
 - The name has been set with numbers and not recipients' names to protect privacy.
 - The last column has the labels for prediction: 1 for spam, 0 for not spam.
 - The remaining 3000 columns are the 3000 most common words in all the emails, after excluding the non-alphabetical characters/words.
 - For each row, the count of each word(column) in that email(row) is stored in the respective cells.

Dataset 2 : Customer Personality Analysis: Analysis of company's ideal customers [\[link\]](#)

- Problem Statement:

Customer Personality Analysis is a detailed analysis of a company's ideal customers and helps a business to modify its product based on its target customers from different types of customer segments.

- Target: Need to perform clustering to summarize customer segments.

- File

- Marketing_campaign.csv, containing several attributes, such as:
 - People: Age, Education, Marital_Status, Income, and so on.
 - Products: Amount spent on different types of product in last 2 years
 - Promotion: several attributes such as promotion campaign or discount
 - Place: the number of purchases made at some place

Dataset 3 : LearnPlatform COVID-19 Impact on Digital Learning [\[link\]](#)

- Problem Statement:
Use digital learning data to analyze the impact of COVID-19 on student learning.
- Target: Need to perform exploratory data analysis to find the impact on Digital Learning under COVID-19.
- File
 - engagement_data: collects page load events of over 10K education technology products in the host's product library.
 - products_info.csv: includes information about the characteristics of the top 372 products with most users in 2020.
 - districts_info.csv: includes information about the characteristics of school districts.

Rule - Proposal submission

- 請於**2021/12/08 (Wed.) 23:59前**在 ceiba 作業區上傳Proposal
 - 組長繳交即可
- Proposal 格式: [[template](#)]
 - 檔名: group<group_id>.pdf (e.g. group1.pdf)

Rule - Presentation slides & short video submission

- 請於 **2021/12/21 (Tue.) 23:59前** 繳交
 - 上傳到 Google drive folder: **"Presentation"**
 - 連結將另寄給各組組長, 擇日公告
 - 組長繳交即可
- Presentation slides
 - 檔名: group<group_id>.pdf (e.g. group1.pdf)
 - 包含: 題目, 組別, 組員
- Short Video
 - 檔名: group<group_id>.mp4 (e.g. group1.mp4)
 - 時長: **8 min.**

Rule - Oral Presentation

- Dates: **12/22 (Wed.) & 12/29 (Wed.)**
- 每組 10 min. (8 min. 報告影片播放 + 2 min. Q&A)
 - 請提早 15 min. 到教室將檔案(short video + slides)放至教室電腦桌面
 - 使用自己筆電的組別也請提前試投影設備

Rule - Report submission & full video submission

- 請於 **2022/01/17 (Mon.) 23:59前** 繳交
 - 上傳到 Google drive folder: **"Report"**
 - 連結將另寄給各組組長, 擇日公告
 - 組長繳交即可
- Report
 - 檔名: group<group_id>.pdf (e.g. group1.pdf)
 - 包含: 題目, 組別, 組員
 - 10~12頁
- Full Video
 - 檔名: group<group_id>.mp4 (e.g. group1.mp4)
 - **20-30 min.**

Thanks for your attention :)

Q & A

- ❖ If you have any questions
 - Email: ntuds2021@gmail.com
 - Ceiba: forum (Discussion on forum is encouraged.)