Project description:

The goal of the project is to apply the Big Data Analytic project cycle (that you studied in lecture) on your selected dataset you like. In the project lifecycle phases, you will use and apply the methods and techniques studied in both lecture and lab. The total marks of the course project will be evaluated according to: (i) project deliverables and (ii) individual oral discussion.

The project steps should meet the following:

1. Search for a real dataset with multiple attributes (a minimum of 10 relevant attributes).
2. Understand the dataset and scientifically define your project objectives w.r.t. data science objectives.
3. Apply data preprocessing/cleaning methods (if needed)
4. Use Hypothesis Testing (if needed).
5. Apply the EDA using different data visualization charts. Select the appropriate chart that allow you to observe more knowledge about your dataset.
6. Mention your comments and observations on your visualizations.
7. Apply any data analytics technique you studied in the course (e.g., ID3, Regression, Apriori, Kmeans, etc.).

Deliverables:

1. Dataset (.CSV file)
2. Code files (R script)
3. Project Documentation [PDF file] that includes:
   a. The project description.
   b. The dataset and variables description.
   c. The problem definition and project objectives.
   d. The data visualization graphics with observations and interpretations of each chart.
   e. Any applied data cleaning or transformation methods (if applied).
   f. Any applied Hypothesis and the test results and interpretations (if applied).
   f. The dataset preparation in terms of machine learning (training set, learning set).
   g. The used data analytics techniques (e.g. ID3, Regression, Apriori, Kmeans, etc.), with justification for your choice.
   h. The performance measures used and the evaluation of analytical technique.
   i. Discussion/Quantification for relevant project findings for your project.

Registration rules and deadlines:

- Project is a teamwork assessment. Same as graduation projects teams.
- Students should register and validate their project dataset and idea with module leader and TAs.
- Link to registration form: https://forms.gle/n2DrcTi42qaA4QVT7
- Each team shall register members' names & IDs, a link to the dataset they will use, project title and short description [140 characters limit].
- Deadline for team's registration: 15-4-2023.
- Project delivery date & discussion will be on the Practical exam date set by the Faculty and CHP administration.