BDAT1007 – Group Final Project

# Due Date: Sunday, April 16, 2023 @ 11.59 PM

# Weight: 40 points of the entire course

# Group Size: 3 – 6 people

## Imagine you (as a team) are running a consulting company specializing in Data Mining. Your potential client varies from small startups to Fortune Global 500 to angel investors and venture capitals.

## Your team is going to present in a global tech conference, and you will showcase some of the hot data mining projects that you have developed. It can be anything like an application, a model, a data pipeline, a workflow, a platform, a service, an API, or a script…

# Deliverables

* Business Cases / Business Requirement Document
* Technical Design Document
* RapidMiner process file (.rmp format) / Source Code
* Raw Data (including the reference links)
* PowerPoint Presentation

# Business Cases

Some questions to consider:

* What’s the purpose of your project? What is the use case?
* Who are your audiences?

Here is a helpful website to seek potential data samples and to enlighten the thoughts.

[Find Open Datasets and Machine Learning Projects | Kaggle](https://www.kaggle.com/datasets)

# Technical Design Document

Some questions to consider:

* What’s the toolset / coding language(s) used?
* How do the data operators & data models look like?
* What does the data flow & data pipeline look like? Any processing mapping / flow chart?
* How large is the data?

# Grading components:

* Originality – WARNING: Plagiarism is a serious academic offence!!!
* Amount of efforts you put on the research and design
* Creativity
* Complexity of the project. Some factors to consider:
  + Data size – we are working with the big data world
  + Data type – are you working with structured data, semi-structured data, unstructured data?
  + Any data preprocessing & transformation?
  + Any data visualization? How do you tell a story from the data?
  + Any machine learning model(s) are used? And why they are used?
* Citations (references). If you reference any sources, you need to clearly put down in a citation section.
* Keep in mind you are doing the project for yourselves, not for me. I hope someday in the future that you can benefit from the research that you did for BDAT1007.

# Submission:

See deliverables. Please also indicate clearly with your group member names and student number.

At least one team member needs to submit on the Blackboard or send me an email.

Please be aware that during the last class (week of April 17, 2022) at least one member from your team is available to present your team project for a few minutes and handle any potential Q & As.