BIF524/CSC463   Data Mining   Fall 2023

Project - Phase 1-

Description: The project is split into three phases that match the learning outcomes throughout the course. Each phase accounts for 10% of your total grade.

Guidelines: The aim of this project is to demonstrate your ability to apply and discuss the outcomes of various data mining techniques on a problem and a dataset of your interest.

- The dataset must include quantitative and qualitative attributes.
- Your work should not be limited to what you learn in the practical sessions of the course.
- You must submit an [R markdown](), knitted as a pdf file, for every phase.
- You can work in a group of two – same group in all phases.
- Your grade will be subject to a 5% penalty for every day of submission delay.

- **Phase I: (10%) due Wednesday, Oct. 20, 11:59pm.**
  - Choose a dataset with a quantitative response, and discuss your choice with me. (1%)
    - **N.B.** Your dataset should not be associated with any existing work related to the required tasks – e.g., on kaggle, Github, …
  - Apply data preprocessing, visualization, and exploration techniques. (4%)
  - Highlight subsequent hypotheses that you can further explore. (1%)
  - Apply linear regression techniques accordingly. (4%)

For each phase, make sure to highlight the following in your R markdown pdf file:

- Dataset description including context and features
- Data mining tasks
- Model performance
- Results
- Comparison of results
- Comments and interpretation

Name of your R markdown pdf file following this template: NameOfTeamMember1-NameOfTeamMember2_Phase PhaseNumber.