

School of Computer Sciences
Universiti Sains Malaysia

Academic Session 2022/2023
Semester 1

CDS501 – Principles & Practices of Data Science & Analytics

Mini Project Guidelines

Overview

The course project is your opportunity to work with existing real data set. Form a group of three (3) members only and register your group at eLearn@USM. Download and explore these datasets and choose your preferred dataset. A dataset is limited to certain number of groups (depending on the total number of groups).

- Drug Review
- Triage
- Parkinsons Telemonitoring
- Customer Review

The project must implement methods that was discussed in class using R language.

Milestones

Deliverable Week	Objective	Deliverable
16 Dec 2022	Assignment 1. Problem Definition and Objective, Dataset Description, Treatment and Exploration (Analysis)	A PDF document
13 Jan 2023	Assignment 2. Choosing and evaluating models to solve problems	A PDF document
27 Jan 2023	Project Report	A PDF document
TBA	Oral Project Presentations (10 minutes per group)	

Submission: via eLearn@USM on the specified dates.

Assignment 1

You need to identify the problem definition and objective, and describe the dataset. You will need to describe step-by-step technique of cleaning and exploring the data. You will also need to report the insight or understanding (conclusion) of the dataset based on your analysis.

You will need to write your report using the given template. The maximum number of pages is **three** (3). You are encouraged to use figures and tables, but the maximum number of both figures and tables is **five** (5) only. The background study should cite **two** (2) to **three** (3) references only.

Use the template as reference.

Assignment 2

You will need to show the process of choosing and evaluating your models to solve business problems identified in Assignment 1. More details will be provided in the class.

Project Report

You will need to write your report in the IEEE paper format max **twelve** (12) pages. You will have the following structure:

- a. Structured Abstract
 - i. Background
 - ii. Aim
 - iii. Methodology
 - iv. Results
 - v. Conclusions
- b. Introduction
- c. Background and Literature Review
- d. Methodology
- e. Results, Model and Discussions
- f. Conclusion and Future Work

Project Presentation

You will need to record the presentation and create an Adobe Spark page. The whole presentation should be 10 minutes long per group. More details about this presentation will be provided during Week 13 class.

Submission Requirements

- All assignments need to be done in group. Submit in groups of four students (maximum). You may choose your own team-mate. Each team submits only **one** report.
- Submission must be made in both hardcopy AND softcopy (submitted online).
- Compress all files (report in pdf and source code) into a zip file. For project, include the presentation slides in pdf in the zip file. The filename **must** follow these naming conventions.
 - <CDS501_Assignment_1_GroupName>
 - <CDS501_Assignment_2_GroupName>
 - <CDS501_Project_GroupName>
- Plagiarism (using other people's ideas and text without proper acknowledgment and using them as your own) is a serious academic offence. The consequences for plagiarism are severe.

.../end