UECS3213 UECS3453 UECS3483 DATA MINING ASSIGNMENT	
Programme(s)	Bachelor Of Science (Honours) Applied Mathematics with Computing Bachelor Of Engineering (Honours) Electronics (Computer Networking) Bachelor Of Science (Honours) Software Engineering
Trimester	January 2023
Course Lecturer	Dr. Fatimah Audah Md. Zaki
Submission Date	18 April 2023 (Week 12)
Submission Platform	WBLE

Student Information					
Group no.	**Student Name**	**Student ID**	**Programme Code**	**Signature**	**Final mark**

Course Learning Outcomes Assessed
CLO2: Create programming solutions using data mining techniques for given problem
CLO3: Evaluate performance of data mining solutions for a given problem
CLO4: Construct a data mining project as a team

Data Mining Assignment: Analyzing a Real-World Dataset

Overview

In this assignment, you will analyse a real-world dataset and apply various data mining techniques to gain insights and answer specific research questions. Your task is to analyse the data, pre-process it, visualize it, answer research questions, and report it.

Dataset

For this assignment, you will be choosing your own dataset to work with. The dataset should be related to a topic of your interest, and it should be publicly available. Follow the instructions below to choose your dataset:

1. Identify a topic of interest. This can be anything from sports to politics to health to social media. Think about what you're interested in and what kind of data might be available on that topic.

2. Search for datasets related to your topic. There are many sources of publicly available datasets that you can use, including:
   - Kaggle: https://www.kaggle.com/datasets
   - UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/index.php
   - Google Dataset Search: https://datasetsearch.research.google.com/
   - Data.gov: https://data.gov/
   - KKMNOW: https://data.moh.gov.my/
   - GitHub: https://github.com/

3. Evaluate the dataset. Once you've found a dataset that you're interested in, take some time to evaluate it. Consider the following questions:
   - Does the dataset have enough data to be useful for your project?
   - Is the data of high quality? Are there any missing values or outliers?
   - Are the data types appropriate for your project?
   - Are there any privacy or ethical concerns with the data?

4. Choose your dataset. Once you've evaluated your options, choose a dataset that is appropriate for your project and that you're interested in working with.

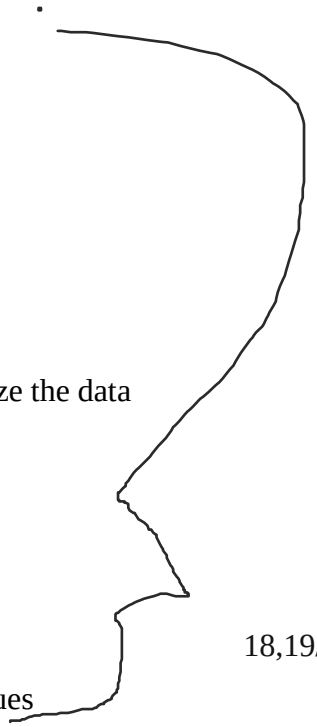
5. Confirm your dataset choice with your lecturer. Before you start working with your dataset, confirm your choice with your lecturer to ensure that it is appropriate for the project and that you'll have the support you need.

Tasks

1. Data Understanding
    - Identify the dataset and describes its source
    - Summarize the variables in the dataset
    - Identifies potential issues or limitations with the dataset
    - Demonstrates understanding of the dataset and its context

2. Data Description
    - Provide descriptive statistics and summaries of the dataset
    - Identify key features, patterns, and trends in the data
    - Use appropriate statistical and visual methods to summarize the data
    - Demonstrate understanding of the data and its properties

3. Data Preprocessing
    - Perform data cleaning, transformation, and normalization
    - Select and justify data sampling techniques
    - Remove any irrelevant data or outliers
    - Demonstrate understanding of data preprocessing techniques



18,19/3

4. Visualization
    - Create effective and informative visualizations to summarize the data
    - Use appropriate chart types and labeling to convey information
    - Provide clear and accurate data labels, titles, and legends
    - Demonstrate understanding of data visualization techniques

1,2/4

5. Research Questions
    - Identify and justify research questions to be addressed using the dataset
    - Develop and apply appropriate data mining methods to answer research questions
    - Evaluate the data mining model to ensure that the model is accurate and reliable
    - Provide meaningful insights and conclusions from the data analysis

8,9/4

6. Presentation
    - Present the analysis in a clear, organized, and engaging manner
    - Use appropriate language, tone, and style for the intended audience
    - Follow appropriate formatting and structure for the presentation
    - Demonstrate good communication skills

15,16/4

Deliverables

Your submission should include the following:

- A Jupyter notebook (or equivalent) with your code, comments, and visualizations.
- A report based on the outlined tasks, summarizing your findings, including a discussion of your approach, key insights, and any limitations or areas for further exploration.
- A brief video presentation (max. 5 minutes) that explains your approach and summarizes your findings.

Marking Rubric

The assignment will be graded out of 120 points; 40 marks for each CLO. The final score will be 30%, which is 10% for each CLO. The marks are based on the following criteria, as outlined in the tasks:

CLO	Criteria	Poor	Satisfactory	Excellent	Marks
2	Data Understanding: **(20 marks)** • Identifies the dataset and describes its source (5m) • Summarizes the variables in the dataset (5m) • Identifies potential issues or limitations with the dataset (5m) • Demonstrates understanding of the dataset and its context (5m)				
	Data Description **(20 marks)** • Provides descriptive statistics and summaries of the dataset (5m) • Identifies key features, patterns, and trends in the data (5m) • Uses appropriate statistical and visual methods to summarize the data (5m) • Demonstrates understanding of the data and its properties (5m)				
3	Data Preprocessing **(20 marks)** • Performs data cleaning, transformation, and normalization (5m) • Selects and justifies data sampling techniques (5m) • Removes any irrelevant data or outliers (5m) • Demonstrates understanding of data preprocessing techniques (5m)				

	Visualization **(20 marks)** • Creates effective and informative visualizations to summarize the data (5m) • Uses appropriate chart types and labeling to convey information (5m) • Provides clear and accurate data labels, titles, and legends (5m) • Demonstrates understanding of data visualization techniques (5m)				
4	Research Questions **(20 marks)** • Identifies and justifies research questions to address using the dataset (5m) • Develops and applies appropriate data mining methods to answer research questions (5m) • Evaluates the data mining model to ensure that the model is accurate and reliable (5m) • Provides meaningful insights and conclusions from the data analysis (5m)				
	Presentation **(20 marks)** • Presents the analysis in a clear, organized, and engaging manner (5m) • Uses appropriate language, tone, and style for the intended audience (5m) • Follows appropriate formatting and structure for the presentation (5m) • Demonstrates good communication skills (5m)				

Note: Poor 0-2, Satisfactory 3, Excellent 4-5

Submission Guidelines

Submit your assignment as a compressed file (zip or tar) that includes the Jupyter notebook, report, and video presentation. Name the file using your group number (e.g., Group1.zip). Submit your assignment through the course's WBLE. The deadline for submission 18 April 2023. Late submissions will not be accepted.