SCHOOL OF COMPUTER SCIENCE

ASSESSMENT TASK 4: Practical (Group Assignment) (Weightage 30%)
JANUARY 2024 SEMESTER

MODULE NAME	: DATA MINING
MODULE CODE	: ITS61504
DUE DATE	: January 23 , 2023, 8.00PM (NST)
PLATFORM	: MyTIMES

This paper consists of SEVEN (7) pages, inclusive of this page.

GROUP NO:

STUDENT DECLARATION

1.   I confirm that I am aware of the University's Regulation Governing Cheating in a University Test and Assignment and of the guidance issued by the School of Computing and IT concerning plagiarism and proper academic practice, and *that the assessed work now submitted is in accordance with this regulation and guidance.*

2.   I understand that, unless already agreed with the School of Computing and IT, assessed work may not be submitted that *has previously been submitted, either in whole or in part, at this or any other institution.*

3.   I recognise that should evidence emerge that my work fails to comply with either of the above declarations, then I may be liable to proceedings under Regulation

No	Student Name	Student ID	Date	Signature	Score

Assessment Questions

The data set we are using in this assignment is the same as the one in assignment 1. You may consider this to be the continuation of assignment 1

Important Notes:

Note 1: *Copying, cheating, attempts to cheat, plagiarism, collusion and any other* attempts *to gain an unfair advantage in assessment result in to award 0 marks to all Parties concerned.*

Note 2: *The Turnitin similarity for this module is 20% overall and lesser than 1% from a single source excluding program source code.*

Note 3: *All the submitted documents will be cross-checked with other students' reports in* this *current and previous semester. Therefore, any similarities rather that whatever is highlighted in Note 2, will be considered as violating assessment rules and a Zero (0) Mark will be given to all group members.*

Note 4: *Severe disciplinary action will be taken against those caught violating assessment* rules *such as colluding, plagiarizing or transcribing.*

Note 5: *The assignment submission document should be within 8 to 15 slides in total with a spacing of 1.5 and a font of 12pt Times New Roman.*

Motivation of Purpose Learning Assignment

Data mining is the process of uncovering patterns and finding anomalies and relationships in large datasets that can be used to make predictions about future trends. The main purpose of data mining is to extract valuable information from available data. It opens many opportunities for businesses, to have a competitive advantage, better understanding of their customers, good oversight of business operations, improved customer acquisition, and new business opportunities.

Your Big Idea and Deliverables

There are two cases study provided below. Students may choose one the case study and require to execute the finding and knowledge based on the course. **Please find the Heart Disease dataset attached on MyTIMES.**

Case Study 1: Predictive Modelling for Heart Disease

Imagine a hospital or healthcare provider aiming to enhance its cardiac care by implementing a predictive model for early detection of heart disease. The dataset includes information about patients who have undergone various cardiac tests. The goal is to develop a machine learning model that can assist healthcare professionals in identifying individuals at a higher risk of heart disease based on their medical attributes.

From the case given, study the given data on heart disease. Credit will be given to those who can produce work beyond the findings and significantly to the final findings.

Case Study 2: Gender-Based Analysis of Heart Disease Risk

A medical research institute is conducting a study to better understand gender-specific risk factors for heart disease. The dataset includes information on patients who have been diagnosed with or without heart disease. Researchers want to explore if there are differences in the risk factors and prevalence of heart disease between male and female patients.

From the case given, study the given data on the depression among male and female. Credit will be given to those who can produce work beyond the findings and significantly to the final findings.

Students are advised to use 1 week for ideation and brainstorming to manage the assignment. Collaborative work with a strong conceptual part is expected.

Assessment Requirements

Students are required to present their idea in the following 2 submission forms:

- Full Report (10-20 pages)

Notes:

Student is not allowed to transcribe directly (cut and paste) any material from another source into their submission.

- Include in-text citation to support your answers and add the list of references at the end of your report (APA format). The list of references is to be alphabetized by the first author's last name, or (if no author is listed) the organization or title.

- The Turnitin similarity for this module is 20% overall and lesser than 1% from a single source excluding program source codes.

Report Components	
Project Title	
Table of Contents	
Abstract (5%)	1. brief introduction of the case study 2. problem statement, aim 3. objective, results
Chapter 1 – Introduction (10%)	1. Project background 2. Problem Statement & Objectives
Chapter 2 – Overview of the datasets (15%)	1. Dataset issues (5%) (Explanation of the data characteristics and source. Include a maximum 1 page of a sample of the actual data.) 2. EDA (10%)
Chapter 2 – Pre-processing Techniques (25%)	1. Observe the issue in the data set 2. Provide justification on the pre-processing used
Chapter 3 – Data Mining Techniques (20%)	1. Develop 3 prediction models 2. Justification on the predictive model chosen

Data Mining Results (10%)	1. Results 2. Performance evaluation of the developed models
Conclusion (5%)	
Submission requirements – 10%	1. Report formatting and pagination 2. Turnitin report, references (APA format)

Submission Requirements

1. Font type: Times New Roman

2. Font size: 12

3. Line spacing: 1.5

4. Alignment: Justify Text

5. Document type: .pdf, .r

6. Number of pages: 10 – 20 pages (do not exceed the page limit)

7. Your full report should consist of the following:

   a) Cover page (Name, ID, Date, Signature, Score)

   b) Marking Rubrics (attach as second page in the report)

   c) Declaration with a clear work breakdown structure to describe what each member is doing (attach as third page in the report)

   d) Report of your answer script

   e) Appendixes (line spacing = 1.0)

   • List of references (APA format)

   • Report of similarity score (percentage of similarity score from each source needs to be shown)

   • Please do not use any AI tools to generate your report. If caught, you will be asked to redo the entire assignment.

8. Your answers/descriptions should be clearly labelled with the questions and sections they are in response to. For clarity, start each question/section on a separate page.

9. All figures and tables are labelled properly.

10. File naming conventions: Group#_ProjectTitle.pdf

Group Project Marking Rubrics (30 Marks)					
Criteria	**Weight (Percentage)**	**Excellent (90-100)**	**Good (75-89)**	**Average (40-74)**	**Poor(0-39)**
Abstract	5	Abstract is written in a clear and concise manner with information to understand the background of the case study.	Abstract is written in a clear manner with some information to understand the background of the case study.	Abstract is written with limited information to understand the background of the case study.	Abstract is poorly written with limited / no information to understand the background of the case study.
Introduction	10	Background is described in a clear and concise manner with information to understand the business nature and example.	Background is described in a clear and concise manner with some information about the business nature.	Background is described in clear manner with little / limited information about the business nature.	Background is described in brief manner with limited / no information about the business manner.
EDA	10	Dataset characteristics are elaborated in very precise and concise manner, with sample dataset given.	Dataset characteristics are elaborated in clear manner, with sample dataset given.	Dataset characteristics are briefly elaborated, with sample dataset given.	Dataset characteristic are listed, with sample dataset given.
Dataset Issue	15	3 dataset issues are described in precise manner with examples given to support the issues mentioned.	2-3 dataset issues are described in precise manner with brief / no examples to support the issues mentioned.	2-3 dataset issues are mentioned with brief / no examples to support the issues mentioned.	1-2 dataset issues are mentioned with brief / no examples to support the issues mentioned.

Data Preprocessing Techniques	15	3 appropriate pre-processing techniques used with well explanation on why the techniques are chosen	2 appropriate pre-processing techniques used with well explanation on why the techniques are chosen	1 appropriate pre-processing technique used with no explanation on why the technique is chosen	Incorrect preprocessing methods applied with no explanation on why the technique is chosen
Data Mining Technique(s) use	20	3 data mining techniques used with excellent parameters justification and explanation on why the techniques is chosen	2 data mining techniques used with good parameters justification and explanation on why the techniques is chosen	1 data mining technique used with well explanation on why the technique is chosen	1 data mining technique used with no explanation on why the technique is chosen
Results (Performance evaluation)	10	Performance evaluation for the 3 data mining techniques with comprehensive explanation	Performance evaluation for the 3 data mining techniques with sufficient explanation	Performance evaluation for the 3 data mining techniques without proper explanation	Performance evaluation for the 3 data mining techniques without explanation.
Conclusion	5	Appropriate conclusion with well explanation	Appropriate conclusion with no explanation	Incorrect conclusion with explanation	Incorrect conclusion with explaination
Submission requirements & team work	10	Fulfill all submission requirements with excellent content organization and team work among members.	Fulfill all submission requirements with consistent content organization and team work among members.	Some effort in compliance with submission requirement and team work among members.	Less obvious effort in compliance with submission requirement and team work among members