CONFIDENTIAL

UTM UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

ALTERNATIVE ASSESSMENT

SCSB4243 SPECIAL TOPICS IN BIOINFORMATICS

SEMESTER 2, SESSION 2021 / 2022

INSTRUCTIONS TO THE STUDENTS:

1. This test book consists of alternative assessment guidelines and rubrics.
2. Please adhere to the guideline upon submitting your technical report.

Fill in your particular below.

Name	
I/C No.	
Year / Course	
Section	
Lecturer Name	

ALTERNATIVE ASSESSMENT DESCRIPTION (40%):

Alternative assessment is assessed **individually**. This assessment is the replacement of your Final Exam. Therefore, failed to submit accordingly, you may fail this course.

By using the obesity patient dataset that was collected from the UCI repository, identify the **efficient classification algorithm** by using Weka 3.8/3.7/Python/R. This dataset consists of 2111 instances with 17 different attributes. This dataset includes data for the estimation of obesity levels in individuals from the countries of Mexico, Peru and Colombia, based on their eating habits and physical condition.

You need to complete the following tasks:

a) Design your proposed method.
b) Pre-processing - data cleaning/ handling noisy data/ discretization of data. Perform at least **TWO (2)** pre-process methods.
c) Machine learning baseline performance using ZeroR algorithm.
d) Learning algorithm - compare your results between at least **THREE (3)** classification algorithms.
e) Evaluate the classification performances and explain the metrics used.
f) Discuss your findings.

Present your completed tasks from the data given and transform the experiment results into a short paper write-up by using the following format:

Short Paper Write up

Write a short paper presenting the workflow and output of the data analysis. Your short paper should consist of:

Chapter	Content
Introduction/Aim	Problem background
Literature	Related works from other researchers
Methodology	Workflow of your method
Finding	Comparison of different algorithms
Discussion	Justification of results
Conclusions	Summary of your experiment
References	List of references

Format: Following the short paper template format as per attached.
Pages: 4 – 6 pages including References.

SUBMISSION:

i. There are **TWO** (2) phases of submission.
ii. Start: **28th June 2022, Tuesday**
iii. Submission:
iv. **Interim submission** - You are required (MUST) to submit your 1st draft of your Python/R program and short write up on **3rd July 2022 (Sunday)** before 12.00am.
    - *Expectation – 50% of overall experiment and write up.*

v. **Final Submission** - You are required (MUST) to submit your FINAL short paper write up on **6th July 2022 (Wednesday)** before 12.00am.
vi. Please submit your data mining short paper write up using the link provided in the e-learning. For your information, different links will be prepared for each submission.

References:

Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in Brief, 104344.

Data Mining Short Paper Write Up Rubric				
Criteria	**1-2**	**3-4**	**5-6**	**7-8**
	SIMPLISTIC	**ACCEPTABLE**	**GOOD**	**EXCELLENT**
Introduction (3%)	Reader cannot determine the problem background OR problem background has no relation to the experiment.	Problem backgrounds are somewhat vague OR problem backgrounds only loosely related to the experiment.	Problem background is fairly clear and matches the writing task.	Problem background is clear to the reader; closely matches the writing task.
Related works (5%)	Offers simplistic, undeveloped, or cryptic support for the related works. Provide no related works to support experiments.	Offers somewhat general support that may be too broad for related works. Provide one to two related works to support experiments.	Offers solid finding but less concrete related works. Provide three to five related works to support experiments.	Substantial, logical, & concrete development of related works. Provide more than five related works to support experiments.
Methodology (10%)	Written with very limited description of steps. Do not include workflow in methodology.	Written with limited description of steps. Include partial workflow in methodology.	Partially clear and not detailed description of steps. Include workflow in methodology.	Clear dan details description of steps. Include workflow of methodology.

Finding (10%)	States very little of what was observed and not in a manner that observations to a conclusion.	No comparison between classification algorithms accuracy is well presented. States very little of what was observed in a manner that observations to a conclusion.	Comparison between classification algorithms accuracy is partially presented. States some of what was observed in a manner that observations to a conclusion.	Comparison between classification algorithms accuracy is well presented. Written in an easy-to-understand manner that ties observations to a conclusion.
Analysis (10%)	No justification from results of comparison between classification algorithms accuracy. Uses no evidence to create a well-supported statement showing knowledge gained.	Comparison between classification algorithms accuracy is very little. Uses very little evidence to create well-supported statements showing knowledge gained.	Comparison between classification algorithms accuracy is not really justified. Uses some evidence to create well-supported statements showing knowledge gained.	Comparison between classification algorithms accuracy is justified. Uses observation, data and other evidence to create well-supported statements showing knowledge gained.
Conclusion (2%)	No discussion on achievement, limitation or contribution. No suggestions for future works.	Limited/minimal discussion on achievement, limitation or contribution. Limited/minimal suggestions for future works.	Somewhat clear discussion on achievement, limitation or contribution. Insufficient or irrelevant suggestions for future works.	Clear discussion on achievement, limitation or contribution. Relevant suggestions for future works.