ASSIGNMENT COVER PAGE

Programme		Course Code and Title	
Bachelor of Information Systems (Hons)/Bachelor of Computer Science (Hons)		CBD3034 Big Data Analysis	
Student's name / student's id		**Lecturer's name**	
		Assoc Prof Dr J. Joshua Thomas	
Date issued	**Submission Deadline**		**Indicative Weighting**
13th October 2023 (Week-7)	5th December 2023 (Week-12)		30%
Assignment [2] title			

This assessment assesses the following course learning outcomes

# as in Course Guide	UOWM KDU Penang University College Learning Outcome
CLO1	
CLO2	
CLO3	Select and apply appropriate intelligent techniques to a given problem.
CLO4	
# as in Course Guide	**University of Lincoln Learning Outcome**
CLO1	
CLO2	
CLO3	Locate and reference relevant information.
CLO4	

Student's declaration

I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.

Student's signature: Submission Date:

Dates and Mechanisms for Assessment Submission and Feedback

Mechanism for handout to students	OL LMS
Mechanism for submission of work by student	*Softcopy online submission via Open Learning*
Date by which work, feedback and marks will be returned to students	7th December 2023
Mechanism for return of assignment work, feedback and marks to students	Feedback will be provided by a marking template. This will be available to students via OL. The discussions at the walkthroughs will also provide informal feedback

COURSEWORK SUBMISSION GENERAL INFORMATION

Academic Integrity Statement

You must adhere to the university college regulations on academic conduct. Formal inquiry proceedings will be instigated if there is any suspicion of plagiarism or any other form of misconduct in your work. Students must **NOT** collude with other groups of students or plagiarize their work.

Nature of the submission required

A softcopy of your assignment in **PDF version** should be submitted to lecturer, no later than the date and time stipulated on the cover sheet. In addition, an electronic copy of your work must be submitted to Turnitin. The first page of your report, immediately after the cover page, must be a page from Turnitin clearly showing your name and your Originality Score (Please refer to submission arrangement).

Diagrams may be used where they are helpful to support your arguments or description. If they are not your own work, the source must be referenced. Please help us to handle and mark your work efficiently.

Documentation guidelines

Student is required to submit a **SOFTCOPY** of the report and ensure that it use the following formatted styles: 1) Font type: **ARIAL**, 2) Font size: **11 pt**., 3) Line spacing: **Single spacing** and 4) Page layouts: **Justify**. Please make sure you have proper format alignment for all paragraphs, following standard writing style and use **HARVARD CITATION STYLE** for citation. Please include a **HEADER** with the following information: **Student ID, Student name, Course code and Assignment type**. Please also include a proper cover page for your submission which contains information about the students, assignment, course, and department with KDU and University of Lincoln (UoL) logos on top. Also include page number and list of references, which is shown in the last page.

Penalties for Late Submission

For late submission of this Assignment, a penalty of a reduction by 10% of the maximum mark may be applicable for each Calendar Day or part thereof that the submission is late. An Assignment submitted more than **TEN** Calendar Days after the deadline will have a mark of zero recorded for this Assignment.

Submission arrangement

1. Cover page
2. Turnitin similarity report
3. Table of Content
4. Main Report
5. Reference List or Bibliography List (whichever applicable)
6. Marking Rubric (in landscape orientation)

This assignment is weighted as **30%** of the assessed work for the course. The assignment is a **group** work.

Students should NOT collude with other students or plagiarize their work. Appropriate action will be taken, according to UOW Malaysia KDU Penang University College regulations, if collusion or plagiarism is suspected. Evidence of academic misconduct will be taken seriously and University College regulations followed. You are advised to be familiar with the University College definitions of plagiarism and collusion. (Refer Handbook) Please do not include references to lecture notes.

You are expected to produce a word-processed answer to this assignment. Please use Arial font with font size of 11.

Background:

The machine learning workbench is a powerful and yet easy to use platform for predictive modeling. You need to equip with your computer or laptop with [1] 3.8.1 or [2] 3.9. You will be given the necessary dataset during the release of the assignment.

Report Guidance

Your report must follow to the below structure and include the required content as described, information on specific marking criteria for each section is available. You must source a written report containing **two** individual sections that provide a full and insightful account of the processes commenced.

Developing a Machine Learning Application

Gene expression data (123 samples) from the microarray experiments of Small Round Blue Cell Tumors (SRBCT) of a Mice cancer study of Khan et al. (2001).

Dataset:
The accompanying data set is taken from 123 cancer samples that were diagnosed by a clinician into four classes (A,B,C,D). For each sample, the expression level of 100 genes has been recorded. Although measuring expression levels is a quick and painless procedure, the full diagnosis requires a painful and lengthy biopsy.

Task A: (Group work -70%)

a) Load the dataset with suitable format (*arff). You need to convert the raw data.

b) You should use more than one data mining algorithms to apply on the given dataset and compare the results.

c) Your solution is able to classify the dataset into the given classes (A, B, C, D) with related diagnosis and explain how the accuracy are measured by the software.

You should explain the process you used to generate your answer, but do not need to give details of how to use the software interface or details of how the classification algorithms work.

Task B: (Individual work – 30%)

You need to write a reflective report (one-page) on what you have learnt by solving Task A. You discussion must focus on assignment solving. This reflective report aims to provide an overview of your experiences and learning throughout the task 1, highlighting key insights and personal involvement. The report adheres to the provided criteria, focusing on the knowledge and understanding gained during the designated timeframe on task allocation materials shared, participating in finding solutions for the questions. Especially through assignment component, enabling a more profound understanding of the subject matter.

[CAI 3024 N Intelligent Systems]						
MARKING RUBRIC **ASSIGNMENT [2]** **(ML Experiment and Report Writing)**						
REPORT COMPONENT (100%)						
LEARNING OUTCOME	**MARKING CRITERIA**	**SCALE**					**YOUR MARKS/COMMENTS**
		Fail (0-49)	**3rd Class (50-59)**	**2nd Lower Class (60-69)**	**2nd Upper Class (70-79)**	**1st Class (80-100)**	
CLO3: Select and apply appropriate intelligent techniques to a given problem.	**1. Introduction (10%)**	Weak or no introduction of the topic. Purpose of the writing is unclear or missing. Topics were not addressed properly.	Basic introduction that states topic but lacks interest.	Adequate introduction and states the topic.	Proficient introduction that is interesting and states topic.	Exceptional introduction that grabs interest of reader and states topic.	
	2. Application of Weka in automating various tasks. (30%)	Fails to show any result of application using WEKA in solving different Computational. No coding is provided.	Partial results of application using WEKA were shown in solving different Computational. Partial coding stage is provided.	Adequate results of application using WEKA were shown in solving different Computational. Adequate coding stages is provided.	Most of the discussions are supported with good elaborations on the data classification with various data splitting methods Most of the coding for Computational are shown.	Good discussions about the data classification with various data splitting methods. Full coding for each Computational ML task is shown.	
	3. Analysis of automation results in term of accuracy and effectiveness of delivering automated task. (20%)	Fails to provide any analyses based on the outputs.	Partial analyses are performed based on the outputs.	Basic analyses are performed based on the outputs.	In-depth analyses are performed based on the outputs.	Complete and in-depth analyses are performed based on the outputs.	
	4. Style of Writing & Citation (10%)	The report writing does not meet the criteria for the assignment (too short or incomplete, too long, and/or completely off-topic). Reference section is missing. Missing or no citation and major flaws on the format.	Many ideas require clarification and/or are off-topic or have marginal relevance to the assignment. Many grammatical and/or spellings errors throughout the paper. The paper is very challenging to read due to poor writing flow. Improper reference section	Ideas are stated clearly and are related to the topic, with only adequate grammatical and/or spelling errors. Reference section with minor flaws. Adequate amount cited works, both text and visual, are done in the correct format. Inconsistencies evident.	Most ideas are stated clearly and are related to the topic, with only minor grammatical and/or spelling errors. Reference section is in minimal. All, both text and visual, are done with minimal errors on the format.	Writing is clear and relevant, with no grammatical and/or spelling errors – polished and professional. Reference section properly formatted. All cited works, both text and visual, are done in the correct format with no errors	

			Very minimal amount of cited works, with incorrect format.				
5. Reflection Report(individual) (30%)		You need to follow the criteria given to you to write the report. No significant study and learning taken place after reading the reflective report.	You report has minimal coverage of the criteria specified	You have explained your reflections significantly, however, there are few criteria in analysis are ignored or not added. No ethical discussions are provided. Overall good	You have precisely explain and covered the criteria provided to you. You have explained the analysis tasks well. However omission of discussion and less structured	You have addressed the criteria given to you and neatly relate the reflective report in terms of technicality, structure, expressing the analysis tasks are clear enough with overall neat reflections.	
						Total (100%)	
						Overall score (100%)	