Universiti Malaysia PAHANG	**COURSE:** MACHINE LEARNING APPLICATIONS	**CODE: BCI 3333**	**MARKS**:
	TOPIC: CHAPTER 2 & 3		
	ASSESSMENT: ASSIGNMENT (INDIVIDUAL)	**DURATION:** 1 WEEK (28/3/2024-5/4/2024)	

Course Outcomes

CO2: Construct appropriate data exploratory and machine learning modeling in a given problem. (Psychomotor: **20%**) (60 Marks)

General Instructions

1. This is an **individual** assignment, and you must use a **Python** programming language.
2. You are required to select **ONE (1)** dataset that is related to either **MOVIE, HEALTHCARE, FINANCIAL, RESTAURANT, GAMES,** or **SPORT** as your case study. Find your dataset of interest from this link: https://archive.ics.uci.edu/ml/index.php and https://www.kaggle.com/.
3. Construct an **Exploratory Data Analysis (EDA)** for your selected dataset to better understand the shape, structure, and pattern of the data, investigate the questions, and develop preliminary insights and hypotheses.
4. Develop a machine learning **REGRESSION MODEL** for your case study to predict a new case.
5. Produce a **report** that consists of the elements below:
   - **a) Description of the case study**
     - i. Provide a complete description of the case study/dataset/topic including references.
   - **b) Exploratory Data Analysis (EDA).**
     - i. Explore the data to find out the descriptive analytics of the dataset to represent the whole structure of the dataset.
     - ii. Identify and handle the missing values in the dataset using **at least TWO (2) methods**. Explain how the missing value is identified and handled.
     - iii. Propose **TWO (2)** graph visualizations that are suitable for your case study. Describe in detail for your graphs.
   - **c) Machine Learning Model.**
     - i. Identify the most appropriate type of classification as your machine learning model for your case study and justify your answer.
     - ii. Design a complete framework of a machine learning model for your case study.
     - iii. Propose **TWO (2)** questions to investigate your machine learning model for your case study and predict the output.
     - iv. Conclude your findings.
6. There is no compromise **(zero mark will be given)** for any kind of **plagiarism** found in your work.
7. You are advised to perform a **backup strategy** such as uploading your document to online storage (SkyDrive, Google Drive, email, free web hosting etc).
8. **Due date submission**: **19TH APRIL 2023, FRIDAY**. Penalty deduct **CO2 - 5% marks** will be given for a late submission.

What you need to submit

1. Pdf Report. Rename your file like this: YourMatricNo-ASSIGNMENT.pdf *(eg: CB20000-Assignment.pdf)*
2. Python source file with output. Rename your file like this: YourMatricNo -ASSIGNMENT.ipynb *(eg: CB20000-Assignment.ipynb)*

Universiti Malaysia PAHANG	**COURSE:** MACHINE LEARNING APPLICATIONS	**CODE: BCI 3333**	**MARKS**:
	TOPIC: CHAPTER 2 & 3		
	ASSESSMENT: ASSIGNMENT (INDIVIDUAL)	**DURATION:** 1 WEEK (28/3/2024-5/4/2024)	

RUBRIC

Category	Descriptor	Level of Achievement						Total mark	Given Marks
		0	1	2	3	4	5		
a. Description of case study	Complete description of the case study/ dataset/ topic including references	Not providing anything	Lack description on case study/dataset. No references provided	<In between>	Partly description on Case study/dataset. At least 3 references	<In between>	Complete and detail description on case study/dataset. More than 5 references	**5**	
b. Exploratory Data Analysis (EDA)	Details descriptive analytics	Not providing anything	Only provide 1 or 2 descriptive analytics	<In between>	Provide descriptive analytics provided, but without explanation	<In between>	Provide details description of descriptive analytics (Features of data, range value used in data, size of the dataset, the sample of data etc.)	**5**	
	Process Identify the missing values	Not providing anything	Not complete process identifying the missing values	<In between>	Provide steps in identifying missing values, but no description	<In between>	Provide steps in identifying missing value with acceptable descriptions	**5**	
	Process of handling the missing values	Not providing anything	Only one method applied for handling missing values	<In between>	Two methods applied for handling missing values, but current description not provided	<In between>	More than two methods were applied for handling the missing values and provided a complete description	**10**	
	Graph visualizations	Not providing anything	Only one Visualization was Produced. No description	<in between>	Two visualizations were produced, but some of the visualizations answered the same questions	<In between>	Two visualizations were produced, covering different aspect in detail	**10**	
c. Machine Learning Model	Type of regression and justification	Not providing anything	Incorrect type of regression. No justification	<in between>	Correct type of regression but incorrect justification	<In between>	Correct type of regression and good explanation for justification	**5**	
	ML framework	Not providing anything	Incorrect framework	<in between>	Incomplete framework	<In between>	Complete and correct framwork	**5**	
	Analysis of the dataset according to the question created	Not providing anything	Only one details analysis provided. No output predicts	<in between>	Provide two details analyses of the dataset, but some of the analyses were overlap. Incomplete output predicts	<in between>	Provide two detailed analyses and hypotheses of the dataset covering each question. Complete predicts	**10**	
	Conclude finding	Not providing anything	Incomplete finding	<in between>	Partly finding explanation	<in between>	Detail finding explanation	**5**	

اونيۏرسيتي مليسيا ڤهڠ

UNIVERSITI MALAYSIA PAHANG

FAKULTI KOMPUTERAN

BCI 333

MACHINE LEARNING APPLICATIONS (ASSIGNMENT)

SESSION 2023/2024 SEMESTER II

LECTURER'S NAME	**:**
NAME	**:**
MATRIC NO	**:**