Assignment 2: *Machine Learning Web Application*

This assessment contributes **30%** of the overall marks for the module **DA4310 Data Science**. This is a **group** assignment of **THREE (3)** students. This assignment is the continuation of the first assignment. Choose only **ONE (1)** dashboard application created in the first assignment to continue in this Assignment 2.

In this assignment, each group is required to create a complete machine learning web application. Students need to create at least three (3) different machine learning models for the dataset, therefore there should be also three (3) different pages in the web application for each model.

The machine learning web application should be able to accept at least 3 different user input. The user input should also be pre-processed before using the model. The machine learning model should be able to output a prediction. The application should also be able to show what features of the data that contributes most to the prediction (if applicable).

Students are also required to pre-process the data before using it for the modelling. Student also need to consider the aesthetic of the application design. Students are also required to continue the previous user manual. The user manual should include step-by-step guide on using the dashboard visualisations and the machine learning models.

On the day of submission, students are required to hand in a softcopy version of the report, source codes and presentation slides to the lecturer. The softcopy versions should be submitted on LMS. Students are required to present their work the week after assignment submission.

Weightage
20%

Due Date
11:59 P.M, 30th March 2024 (Saturday)

Late Submission
Deliverables received after 11:59 P.M. will result in 10% mark deduction per working day.

Deliverables:
Report Softcopy, source codes and Presentation Slides. Submission through LMS.

Module Lecturer
Norfarrah Muhd Masdi
norfarrah.muhdmasdi@pb.edu.bn

Resource
Students own research
Lecture slides & DSCI portal.

Equipment/Software
Students are required to use their own personal computer and necessary software.

Task Checklist

Please read the following for a detailed explanation regarding what you need to do to complete this assignment. If you have any questions regarding the content, kindly request assistance from your respective module lecturers for further understanding. You are required to submit a **softcopy version** of the **report, source code** and **presentation slide**.

1. Report

The cover page of this report should contain the template provided on the LMS. The contents should follow the following format:

- Table of content
- Documentation:
  - **Introduction** - Introduction on the topic. Determine the problem statements, aims and objectives, and the purpose of the application.
    - **Problem Statements** – Determine the problem statements. The problem statements should be supported by clear and valid evidences.
    - **Aims and Objectives** – Identify the aims of the dashboard application. Determine at least 3 objectives.
    - **Application Overview** – Describe the application and the purpose of the application.
  - **Machine Learning Models –** Screenshot of each machine learning model's interface. Descriptions of each machine learning model page. Discuss the importance of each model in achieving the objectives. This should also include the important features extracted from the model along with its descriptions. Provide your inference based on the features extracted.
  - **User Manual** – Brief description of the application's features. Create a step-by-step guide for using the application.
  - **Conclusions** – Summary of important points from your report. Discussion on how the application achieve the aims and objectives.
- References - Table (uses APA format and contains resources used as references in this report)

2. Application Minimum Requirements

The application must contain but not limited to the following:

- Data Preparations
  - This should contain all data cleaning and transformation techniques applied to the dataset(s). Only data cleaning and transformation of data that is use for creating the model will be marked in this assignment.
- Interactive Main Dashboard:
  - Main Dashboard which must contain at least 6 visualisations

- All visualisations must have filter functions where the visualisation changes based on the user input.

- Machine Learning Model Pages:
  - At least 3 machine learning models implemented.
  - At least 3 different input type in each page.
  - All input should be pre-processed before prediction.
  - Constraints applied to user input wherever applicable (example: phone number input can only accept number values).
  - Output prediction based on user input.
  - Each page should also contain the details of the model such as the name of model use and the result of the model evaluation.
  - Each page should include the important features extracted from the model (if applicable).
- Libraries List:
  - This should contain all libraries used/installed within the application.
- Application Design:
  - The application must be well-designed and can be easily accessed with proper colour scheme, layout and usability factors.
  - Marks will be rewarded for aesthetics of the application
  - Templates are allowed but templates must be cited and referenced.

3. Demonstration

As a group, you need to include the following contents in your presentation slides. Your presentation slides should address the following:

a) Introduction
b) Overview of the application
c) Demo of the application
d) Conclusion

General Guidelines:

- Presenters will have **a maximum of 20 minutes** to deliver their presentation.
- The presentation should address the contents specified in the previous section.
- Lesser and shorter point is better and fill in the PowerPoint slides with narration, not words.
- A question-and-answer session will be at the end of the presentation.

Assessment Rules

1. Plagiarism is a serious academic offence. You will be penalized heavily if caught plagiarizing. You must take necessary steps in ensuring your work is plagiarism-free. **For severe cases of plagiarism, grade 'F' (Fail) will be awarded. Zero marks will be recorded for your first attempt. The same applies to collusion.**

2. The use **of AI-tools** or any other type of automated assistance is strictly prohibited in completing the assignment. You will be penalized heavily if caught. You must take necessary steps in ensuring your work is AI-free. If caught, you may receive **grade 'F' (Fail)** and disciplinary actions may be taken.

3. If you have problems in completing the assessment due to illness, you must **REPORT** to the lecturer immediately. Last minute notification will not be entertained unless student presents a medical report, which is issued and endorsed by any government health centres.

4. You may consult with the module lecturer if you are not clear with the assignment by approaching via e-mail on initial. Module lecturer has the right not to entertain any questions except for questions in regards to a submission within three (3) working days before the assignment's deadline.

5. In the event you are unable to complete the assessment or submit partially completed assessment, the following actions will be taken:

    a. For **significantly incomplete** assessments or **non-submission**, you will be automatically awarded with grade 'F' (Fail). You will be asked to complete the assessment however as a **second attempt**. Marks will be capped at 50%.

    b. For submission of **partially completed** assessment, marks will be awarded as per work submitted. Should marks fall below the passing grade, you will be referred to case 3a above.

6. It is your responsibility to check the LMS regularly for any new updates or announcement. Do not simply rely on your friends to get updates.

7. Rules may be updated from time to time. Any updates will be posted on the LMS.

Report Writing Instructions

- Presentation of report including spelling, grammar, syntax and style will be marked.
- Please remember that your report should be formatted according to the **APA Style Referencing guide** and follow all rules stated in the criteria section.
- Your report must be a word processing document including cover page, table of content, references and appendices.
- Cover page must include the module code, module name, title of the assignment, semester and academic session, student's full name, student ID code, programme title, submission date, module lecturer full name and Politeknik Brunei logo only.
- Use of **12-POINT FONT**, the **ARIAL** or **TIMES NEW ROMAN** font, single or **1.5** spaces, and **PAGINATE**.
- Penalty of marks will be applied if it does not follow this specific format. Kindly note that the cover page must be pasted on top of the folder.
- Assessment must be submitted by the due date and time shown. **Late assessment will be penalized to 10% deduction of the total possible marks for the assessment for every working day after the deadline.** Where the assessment is submitted more than one week late, a mark of zero shall be awarded.
- **It is the student's responsibility to ensure no aspect of their work is plagiarized or the result of other unfair means.**

Deliverables

The list below is the deliverable each group MUST submit. In the case that deliverables received on the submission date after 11:59 P.M., marks will be deducted by 10% each day and will continue to reduce accordingly. More than 10 days would result in automatic failure and will be given a second attempt for this assignment with barring marks.

- Report Document (in PDF format)
- Source code – in python script format (.py)
- Presentation Slides (in PDF format – 1 slide per page)

<u>Important Note:</u>

1. Document should be renamed according to the format: DSCI-AS2-StudentId-Name01-Studentid-Name02-Studentid-name03.pdf, Example: DSCI-AS2-19FTT2020-NORFARRAH-19FTT2021-NORFARRAH-19FTT2022-NORFARRAH.pdf
2. Only one (1) PDF file is needed for PBLMS online submission.
3. Students are advised to check on the document before submission. A penalty of up to 30% marks deduction for submission of corrupted files.
4. No hardcopy submission is required for this assignment.

5. In the event of technical error in submission, the module lecturer will contact you to resubmit. The corrected version MUST be submitted within one (01) working day. Students are encouraged to keep a copy for resubmission ready.
6. Although the submission date is on **30th March 2024**, students are strongly advised to consult your lecturer on weekly basis to keep you on the right track and detect any errors so amendments and corrections can be made at early stage before furthering the work.

   Lecturer is there to guide students, not to provide the solution. Some of the contents require for this assignment will not be available in the early lectures so students are expected to use their own time to do research in advance and to make sure they are on schedule.

Assignment Consultation

Assignment consultation is compulsory for each group. Students must schedule a consultation date and time with the module lecturer. It is recommended for the students to schedule a consultation in Week 7. The following dates are recommended for the consultation:

1. 09th March 2024
2. 10th March 2024

Important Note:

1. Students are expected to demonstrate satisfactory progress during the consultation. In instances where there is minimal progress, additional consultation sessions will be required to address any challenges or concerns.
2. Prior to attending the consultations, students are to prepare their questions before attending the consultation.
3. Failure to schedule and attend the consultation will including missing the additional consultation result in a 5% deduction from their overall marks.

Grading Criteria

Table below is the guideline on how this assessment will be graded.

Task	100%	Guide to score full marks
Report (25%)		
Report Structure	1	Proper cover page as instructed in this assignment brief. Includes: table of content, proper font, font size, spacing and page numbering.
Introduction	4	• Introduction on the topic

		• Problem Statements – Determine the problem statements. The problem statements should be supported by clear and valid evidences. • Aims – Identify the aims of the dashboard application. Only one aim required. • Objectives – Determine at least 3 valid and measurable objectives. • Application Overview – Describe the application and the purpose of the application.
Machine Learning Models	8	• Screenshot of each machine learning model's interface. • Descriptions of each machine learning model page. • Discuss the importance of each model in achieving the objectives. • Important features extracted from the model along with its descriptions. • Provide your inference based on the features extracted. At least one inference from each model
User Manual	8	• Brief description of the application's unique features. • Create a step-by-step guide for using the application (only unique features to avoid repetition). • Include screenshots wherever necessary.
Conclusions	3	• Summary of important points from your report. • Discussion on how the application achieve the aims and objectives.
Reference	1	Proper indication on the text/paragraph where a concept, explanation or solution are taken or reference from. A table of references following APA standard referencing.
Application Implementation (50%)		
Data Preparation	5	• Suitable data preparation technique used on the dataset for modelling. • Data preparation's steps are done correctly for modelling.

Data Visualisation	5	• Data visualisations are done properly • Suitable visualisation is used to present the data. • Visualisation conveys correct information. • Different types of visualisations are used. • Visualisation's aesthetic (visualisations are labelled properly, consistent format and clear title) • At least 2 different types of user interactions used. • All visualisations changed based on user input.
Machine Learning Models Implementation	30	• At least 3 machine learning models implemented correctly. • At least 3 different input type in each page. • All input should be pre-processed before predicting the output. • Constraints applied to user input wherever applicable (example: phone number input can only accept number values). • Correct output prediction based on user input. • Important model details are captured. • At least 2 model evaluations technique used correctly. • Important features are captured and shown in the page.
Application Design	10	• The overall interface design promotes readability, completeness, great visual presentation and quality work done. • Transitions across pages are smooth and clear. • All features work as intended.
Demonstration (25%)		
Presentation	5	• Excellent presentation skills including communicating ideas, voice projection and body language.
Code Explanation	4	• Able to explain the process and code used to develop the applications.

User Interface Design and Experience	3	• Show a well-designed user interface that enhances user experience.
Execution and Functionality	5	• Demonstrate functional models' prediction. • Demonstrate dashboard functionalities with effective insights.
Data Insights and Predictions	3	• Ability to produce meaningful insights. • Ability to produce meaningful predictions.
Questions and Answers	5	• Ability to answer questions with accuracy.

- End of File –