Assignment 2: *Applied Supervised/Unsupervised Learning*

This assessment contributes **30%** of the overall marks for the module **DA3304 APPLIED PROGRAMMING FOR DATA ANALYTICS.** This is a **group** assignment of **THREE (3)** students with the possibility of one group ending with **TWO (2)** total members only. Each group is required to choose at least **one (1)** dataset.

Dataset can be obtained from but not limited to the following:

- Google Dataset Search https://datasetsearch.research.google.com/
- UCI Machine Learning Repository archive.ics.uci.edu/ml
- Open collaborative machine learning platform www.openml.org
- Kaggle: Platform for competitions, tutorials, datasets, jobs etc. www.kaggle.com

Each group is not allowed to choose the same datasets for this assignment. Students are not allowed to use any datasets that are used during the practical session. Each group is required to inform the lecturer on which datasets are chosen to avoid any clashes.

Student's task is to create different models using supervised or unsupervised learning and compare the models using appropriate model evaluations. Students are also required to develop meaningful visualisations that describes the chosen dataset. Students are to go through data analysis processes starting from business understandings to model evaluation. Students are required to documents and to produce detailed report on steps performed to achieve the result.

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 the week after assignment submission.

Weightage
30%

Due Date
11:59 P.M, 16th April 2022 (Saturday)

Late Submission
Deliverables received after 11:59 P.M. will result in the deduction of 10% each day and will continue to reduce accordingly.

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 & APDA 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. The breakdown of the report's content.
  - **Business Understanding** – Determine the source(s) of the data. Describe the industry/business associated with the dataset. Identify possible business benefits from the datasets. Formulate business objectives in relation to the dataset.
  - **Data Understanding** – Identify the chosen dataset. Describe the attributes and identify the dependent and independent variables. Suggest with justification different attributes that may contribute to data analysis.
  - **Data Preparation** – Identify noisy or/and dirty data. Describe the process taken to clean and format those data. Explain the need to clean or transform those data. Includes appropriate tables that document the changes made to the dataset.
  - **Data Visualisation** – Explore the dataset and provide appropriate visualization. Describe the results of exploring the dataset and highlight any interesting information and patterns discovered.
  - **Modelling and Model Evaluation** – Determine the learning algorithms chosen for the dataset. Applied the chosen algorithms and record the outcome. Evaluate different models and any variations of the models.
  - **Recommendations** – Recommend the best model and justify your recommendations. Determine ways that can improve your model that can help your businesses.
  - **Conclusions** – Summary of important points from your report. Discussion on how the model helps fulfill the business objectives.
- References - Table (uses APA format and contains resources used as references in this report)

2. Jupyter Notebook / Source Code

Your source code must contain the following:

1. Data Preparation – This should contain all data cleaning and transformation techniques applied to the dataset(s).

2. Data Visualisation – This should contain all visualisations created.
3. Modelling and Model Evaluation – This should contain all models with its variations. It also should include the model evaluation result of the model.

3. Presentation

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

a) Introduction
b) Business Understanding
c) Data Visualisations
d) Models and Models Evaluation
e) Recommendation
f) 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. 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.

3. 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.

4. 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.

5. 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 Jupyter Notebook format (.ipynb) or python script format (.py)
- Presentation Slides (in PDF format – 1 slide per page)

Important Note:

1. Document should be renamed according to the format: APDA-AS2-StudentId1-StudentId2-StudentId3.pdf, Example: APDA-AS2-19FTT2020-19FTT2021-19FTT2022.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 **16th April 2022**, 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.

Grading Criteria

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

Task	100%	Guide to score full marks
Report Structure	2	Proper cover page as instructed in this assignment brief. Includes: table of content, proper font, font size, spacing and page numbering.
Introduction	3	• Introduction on the topic • Breakdown of report
Business Understanding	7	• Determine the source(s) of the data. • Describe the industry/business associated with the dataset. • Identify possible business benefits from the datasets. • Formulate business objectives in relation to the dataset.
Data Understanding	8	• Identify the chosen dataset. • Describe each attribute in the dataset • Identify the dependent and independent variables. • Identify the type of data for each attribute. • Suggest with justification different attributes that may contribute to data analysis.
Data Preparation	10	• Identify noisy or/and dirty data. • Describe the process taken to clean and format those data (include necessary tables and/or figures). • Explain the need to clean or transform those data. • Data preparation's steps are done correctly (source code)
Data Visualisations	15	• At least 10 visualisations are included. • Descriptions of each visualisation. • Data visualisations are done properly (source code)

		• Visualisation are easy to understand (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)
Modelling and Model Evaluation	35	• Describe the chosen algorithms. • Describe at least 2 model evaluations methods. • Determine at least 2 learning algorithms chosen for the dataset. • At least 2 variations of each algorithms applied. • At least 2 model evaluations technique used. • Describe the parameters changed to create models. • Accurate steps implemented for modelling. • Accurate steps implemented for model evaluation. • Results are properly recorded (report and source code).
Recommendations	8	• Identify at least 2 strengths and 2 weaknesses of your models. • Suggest at least 2 ways to improve to any models created.
Conclusions	5	• Summary of important points from your report. • Discussion on how the model help fulfil the business objectives.
Reference	2	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.
Presentation	5	• All sections are covered in the presentation according to the guidelines. • Contents are organized and easy to follow. • Clear and strong voice projection and appropriate body language projection. • Able to explain all of the features and slides without difficulty.

		• Able to answer all questions posed by the audience without difficulty.

- End of File –