

## ICT304 Tutorial 3

**Due: Teaching Week 7**– Submit the tutorial to your tutor during your tutorial session in week 7. External students will need to upload the answers to LMS and an interview session may be arranged by your tutor.

**Note:**

- You must provide answers and discussions in a PowerPoint file, which must be self-explanatory. You may be asked by your lecturer/tutor to present your answers during your tutorial session.
- You need to provide references to the sources used to answer your questions. Visit the library website to learn how to do proper referencing. See <https://www.murdoch.edu.au/library/support/referencing-guides>
- If the question requires you to describe the installation steps, the code, or the program, you may also need to demonstrate them to your lecturer/tutor.
- **You are discouraged from using AI tools like ChatGPT to answer your questions, and if you use such tools, you must declare or reference the part that you use them. See <https://libguides.murdoch.edu.au/APA/generativeAI> for information on how to make reference to it.**

**Background:**

Before attempting the following questions, you should

- Read the lecture materials for Week 5 and Week 6
- Read the references provided in the lecture slide for Week 5 and Week 6

**Tutorials:**

- 1) Explain the differences between the different type of samplings. **Illustrate with examples** the advantages and disadvantages of each type of sampling methods.
- 2) For the dataset below, it has the problem of imbalanced data. Provide the steps and codes for dealing with the problems before building a Machine Learning model. Provide a comparison study of the performance of the Machine Learning model **before and after** the imbalanced data issue has been handled.

Abalone Data Set: <https://archive.ics.uci.edu/ml/datasets/abalone>

- 3) Explain the differences between batch processing and stream processing **using examples**. Discuss at least two methods of Machine Learning techniques that can be used for stream processing.

- 4) Provide the discussion of your research required in the lecture notes. Prepare a 10mins PowerPoint presentation to explain the following, you may be asked to present this during the tutorial class.
- How is feature importance relating to feature selection?
  - Do you know the different approaches of performing feature extraction?
- 5) Complete the tutorial on data leakage in the following URL. After the completion of the tutorial, prepare a 5-mins PowerPoint presentation to discuss the problem of data leakage, and what you have learned from this tutorial.

Data Leakage: Find and fix this problem that ruins your model in subtle ways

<https://www.kaggle.com/code/alexisbcook/data-leakage/tutorial>