# Overview

During this assessment you will produce a written report on analyses of a real-world problem using a neural network and an alternative machine learning method. For both techniques, you will investigate the properties of the methodologies by altering the machine learning parameters. Additionally, you will compare and contrast the two methodologies using the same data source.

The purpose of the project is to build AI & ML capability to:

* apply machine learning methodologies; undertake independent research
* investigating machine learning parameters; compare and contrast machine learning methodologies;
* construct a written communication and interpretation of findings resulting from machine learning methodologies.

# Submission

**Report Structure**

The report **MUST** follow the below structure.

* Title and author
* Abstract/Executive Summary (one page)
* Brief overview on the Australian unemployment rate
* Data. Including any data preparations and sub-setting
* Analysis and Investigation of Machine learning (ML) method
* Analysis and Investigation neural network (NN) method
* Comparison and Contrasting ML and NN models
* Conclusions/findings, together with any recommendations and lessons learned
* Bibliography/References - not included in page count
* Appendices, (e.g. Rcode, additional graphs, keys and definitions) - not included in page count

**Submission**

You will need to submit the following:

1. A PDF/Word file. The assignment can be a **maximum of 12-A4 pages**. References and Appendices are not included in page count.
2. **Rmarkdown/R** script file to reproduce your work.

**A word on plagiarism:**

Plagiarism is the act of using another’s words, works or ideas from any source as one’s own. Project work containing plagiarised material will be subject to formal penalty processes.

# Business scenario

## Dataset

The data, **“AUS data 2023.xlsx”**, used in the project is aggregated and collected from the Australian Bureau of Statistics (ABS).The data is available quarterly from Dec 1982 to March 2023. The data includes the response variable (unemployment rate) and 7 predictors:

* *Y*: unemployment rate measured in percentage
* *X*1: Percentage change in Gross domestic product;
* *X*2: Percentage change in the Government final consumption expenditure;
* *X*3: Percentage change in final consumption expenditure of all industry sectors;
* *X*4: Term of trade index (percentage)
* *X*5: Consumer Price Index of all groups (CPI) ;
* *X*6: Number of job vacancies measured in thousands;
* *X*7: Estimated Resident Population in thousands.

# Assessment Tasks

1. Use the data from Dec 1982 to Dec 2020 as the training set
2. Provide an overview of the Australian unemployment rate over the training data period, and some insights on factors driving the unemployment rate (provide relevant references when needed; maximum one A4 page).

## Data

1. Prepare data appropriate for the proposed supervised machine learning methodologies such as:
2. implementing appropriate data wrangling procedures, e.g. missing values treatment /transformation of variables.
3. provide and comment on descriptive statistics of the variables.

## Machine Learning

4. Select the most effective supervised machine learning (ML) algorithm to the dataset prepared in Question 3 to predict the Australian unemployment rate from March 2021 to March 2023.

1. Justify your choice over the other supervised machine learning algorithms.
2. Justify the choice of the hyper-parameter(s) which is required to be specified in R to estimate the selected model.
3. Report the performance(s) and interpretation(s) of the obtained ML model(s) on the training dataset.
4. Discuss the predictive performance of the model on the test dataset (March 2021 to March 2023).

## Neural Network

5. Apply a neural network (NN) to the data prepared in Question 3 to predict the Australian unemployment rate from March 2021 to March 2023.

1. Describe the structure of the selected neural network model.
2. Report the performance(s) and interpretation(s) of the produced NN models on the training dataset.
3. Discuss the predictive performance of the model on the test dataset (March 2021 to March 2023).
4. Vary the number of hidden layers in the model 5(a). Explore the impacts of the change on the prediction performance of the model.
5. Vary the number of neurons in each layer in the model 5(a). Explore the impacts of the change on the prediction performance of the model.

## Comparison and Suggestion

1. Compare the chosen ML model in Question 4 with the NN model in Question 5, and then provide a recommended model. At a minimum, include
   1. Cross-validated accuracy
   2. Computational time to train models
   3. Interpretability
2. Provide some suggestions regarding the methodologies/data to further improve the prediction of the unemployment rate of Australia.