# Final Assignment (40%)

# Release Date: 17th Dec 2020

# Submission Date: 15th January 2021

# Machine Learning

# Higher Diploma in Artificial Intelligence/ Data Analytics

**Instructions for this final assignment**

- This is an open book assignment under the following guidelines.
- You are allowed to use Lecture, Tutorial notes and online internet material, but the copy of the text is not allowed.
- Turnitin database will be used to check the plagiarism of this assignment.
- The final assignment will be uploaded at your Moodle page.
- If you are using pen and paper for this assignment. After completion of your work, take a picture with your mobile phone camera then you submit the picture on Moodle. Please ensure that you write below your drawing the question number the drawing corresponds to.
- Use a single column layout document.
- Font size for the body of the text should be 12 point Arial/ Calibri.
- Write student ID, Name, Module Name and undergraduate year at the top of the first page
- Clearly write the question number in the assignment.
- Answer the questions based on the specified number of words.
- Harvard Reference Style will be used

Attempt all questions and the marks for each question are mentioned separately.

## Question 1:

What is the difference between supervised, semi supervised and Reinforcement machine learning approaches? Briefly discuss and describe using a diagram and with examples.

(300 - 350 words, 15 marks)

## Question 2:

You are given a Boeing aeroplane data set having 850 columns and 2 million rows. The data set is based on a classification problem. Your project director asked you to reduce the dimension of this data so that the computation time for machine learning model could be reduced. What method would you suggest to resolve the above-mentioned problem? (You may consider any practical assumptions.)

(300 - 350 word, 15 marks)

## Question 3:

a) What are convolutional neural networks? Where can we use them? Describe three kinds of activation functions that are considered as effective to solve neural network problems.

b) Demonstrate an importance of Cross-Validation for the authentication of Machine Learning Modelling outcomes. Briefly clarify and explain the strategy for the implementation of Cross Validation technique for Machine Learning models.

(300 – 400 words, 10 + 10 = 20 marks)

## Question 4:

a) Briefly discuss and illustrate the Decision tree approach for classification in Machine learning. If you are running a binary classification tree algorithm. Do you know how does a tree splitting takes place i.e. how does the tree decide which variable to split at the root node and succeeding nodes?

b) Describe the role of Underfitting and Overfittings the models in Machine Learning. Briefly explain and suggest steps how a data scientist can avoid these problems.

(400 - 500 word, 15 + 10 = 25 marks)

## Question 5:

a) What is the purpose of Market Basket Analysis. How the association rules mining benefit the manager to perform better decision making to satisfy the needs for the customers as well as an organization?

b) What are the major goals of clustering analysis? How is kNN different from kMeans clustering? Briefly illustrate and interpret by providing an example.

(400 - 500 word, 15 + 10 = 25 marks)