Task 1 P, C: ML Development Using Python

This document supplies detailed information on Assessment Task 1 for this unit.

Key information

- **Due: 10th March 2024, by 11.59 pm IST.**

Overview:

In week 1, you have explored the different machine learning algorithms (supervised and unsupervised learning models) and how you can develop a decision tree from scratch using Python programming. This will help you to understand how to build an ML model using Python and then ready to move on with creating other Models.

- **The students who target Pass:** For this task you need to create a ML model based on a decision tree and submit the code along with the screen shot that you got it working. You need to provide the explanation of the selected dataset, the developed code line by line and discuss the output and the performance of the built model using the selected dataset. **You need to zip the model at the end** (Pass task).

- **The students who target Credit or Higher:** For this task you need to develop two ML models based on Decision tree and Random Forest and provide the codes for both models that you got them running and compare the models in terms of accuracy and some other metrics and provide justification which model is performing better and why. Please keep in mind that you do not need to export and zip two models. **You need to export and zip the best model in your code.**

The ML model should be built using a dataset. You need to look up a dataset from the internet. I will suggest the following datasets:

https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease

To do this assignment, you need to refer to Week 1 content.

Submission details to Olympus:

- **Submit your answers as a PDF file to Olympus. Your answers must be relevant and precise. In your submission you need to tell us your target grade.**
- **In your submission you need to answer the following parts:**

1. What is the selected dataset and what is the related problem for this dataset? You need to provide details of datasets, dataset description, what are the features, output (class label) and discuss the problem that needs to be solved by machine learning model. (Minimum 200 words)

2. You need to provide the screenshot of the built ML pipeline (Data ingestion, Data preparation, model training and evaluating the model). You need to provide a cell by cell explanation of the code.

3. What is the performance of the build model/ models (Based on your target grade)? You need to provide discussion and justification of how the model is performing (discuss different metrics like accuracy, confusion matrix, etc.) based on the selected dataset.

This part if for students who target **Credit or higher**:

4. You need to compare the performance of the models (using different metrics) and provide justifications which mode is performing better and why.

Academic integrity, plagiarism and collusion

Plagiarism and collusion constitute extremely serious breaches of academic integrity. They are forms of cheating, and severe penalties are associated with them, including cancellation of marks for a specific assignment, for a specific unit or even exclusion from the course. If you are ever in doubt about how to properly use and cite a source of information refer to the referencing site above.

Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgement as to its authorship, the work of any other person or resubmits their own work from a previous assessment task.

Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose, with the intent of obtaining an advantage in submitting an assignment or other work.

Work submitted may be reproduced and/or communicated by the university for the purpose of assuring academic integrity of submissions:
https://www.deakin.edu.au/students/study-support/referencing/academic-integrity.