Task3.2: ML development using Python

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

Key information

- Due: **Monday 10 April 2022** by 8.00 pm (AEST)
- **This assignment is graded (from Pass to Distinction), and you can submit this assignment only once and there is no chance of resubmission.**

Overview:

During week 2, 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 and credit: 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 your 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 higher than Credit (D): For this task you need to create 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 (Distinction). 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 dataset:

https://archive.ics.uci.edu/ml/datasets/Acute+Inflammations.

To do this assignment, you need to refer to Week2 seminar presentations.

Submission details to OnTrack:

- **Submit your answers as a PDF file into the Ontrack. 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 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 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 explanation for cell by cell 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 matrices, accuracy confusion matrix) based on the selected dataset.

This part if for students who target **higher than credit**:

4. you need to compare the performance of the models and provide justifications which mode is performing better and why.

Submission details

Deakin University has a strict standard on plagiarism as a part of Academic Integrity. To avoid any issues with plagiarism, students are strongly encouraged to run the similarity check with the Turnitin system, which is available through Unistart. A Similarity score MUST NOT exceed 39% in any case. **No marking on any submission after due date.**

Extension requests

Requests for extensions should be made to Unit/Campus Chairs well in advance of the assessment due date. If you wish to seek an extension for an assignment, you will need to submit a request using the OnTrack system as soon as you become aware that you will have difficulty in meeting the scheduled deadline, but at least 3 days before the due date. When you make your request, you must include appropriate documentation (medical certificate, death notice) and a copy of your draft assignment. Conditions under which an extension will normally be approved include:

Medical To cover medical conditions of a serious nature, e.g., hospitalisation, serious injury or chronic illness. Note: Temporary minor ailments such as headaches, colds and minor gastric upsets are not serious medical conditions and are unlikely to be accepted. However, serious cases of these may be considered.

Compassionate e.g. death of close family member, significant family and relationship problems.

Hardship/Trauma e.g., sudden loss or gain of employment, severe disruption to domestic arrangements, victim of crime. Note: Misreading the timetable, exam anxiety or returning home will not be accepted as grounds for consideration.

Special consideration

You may be eligible for special consideration if circumstances beyond your control prevent you from undertaking or completing an assessment task at the scheduled time. See the following link for advice on the application process: http://www.deakin.edu.au/students/studying/assessment-and-results/special-consideration

Assessment feedback

The results with comments will be released within 5 business days from the due date.

Referencing

You must correctly use the Harvard method in this assessment. See the Deakin referencing guide.

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.