INT104 – Artificial Intelligence

Coursework 2

Introduction

In this coursework, a spreadsheet has been provided to perform a set of data analysis. The spreadsheet contains the following information: the index of student, gender of student, the programme that a student is enrolled, the grade that the student is in, total marks that a student is awarded and the mark of 5 exam questions (indexed as MCQ, Q1, Q2, Q3, Q4 and Q5).

The index student ranges from 1 to 619. The gender of the student is represented as "1" and "2". The grade of the student is either "2" or "3". The programme of the student is represented as "1", "2", "3" and "4". The full mark of the whole exam paper is 100. The full mark for 5 exam questions are 54 marks (MCQ), 8 marks (Q1), 8 marks (Q2), 14 marks (Q3), 10 marks (Q4) and 6 marks (Q5) respectively.

The coursework requires students to extracts features of the data and analyse the distribution of the feature with association of the programme that a student is enrolled.

Tasks

1. Use decision trees to classify the programme that a student comes from. Comment on the different decision tree resulted. You may want to try random forest and compare the result. You may also try different features to pursue the best result.
2. Classify the programme that a student comes from with SVM (Support Vector Machine). Try the best configuration of SVM with your preferred features.
3. Classify the programme that a student comes from with Naïve Bayes. Try different sets of features. Please explain how the features impact the classification process.
4. Build up an ensemble classifier. Compare the choice of feature that associates with different classifiers.

Requirement

You MUST use Python to perform the dedicated tasks. Over the lab session, a Teaching Assistant (TA) will be assigned to you to support your work. When you finished your experiment, please make sure you have demonstrated your work lively (in order to make sure the experiment is designed and performed by you). However, please bear in mind that it is NOT the TA's responsibility to teach you Python programming nor design the experiment for you.

After the lab session, you should write a lab report that documents the experiment you performed, the results you obtained and the discussion that justifies your recommended way to classify the provided dataset.

A MATLAB script that guides the experiment has been provided separately. The student is kindly reminded that 1) implementing the provided MATLAB script with MATLAB will result a mark of zero; 2) implementing the provided MATLAB script with Python does not guarantee a high mark.

Over the live demonstration, you will be asked for no more than three questions related to 1) your code, 2) the algorithm of you have used, and 3) the results you have obtained. You may also need to make minor changes to your Python script upon request and explain the corresponding results.

The length of the lab report after the lab session should be no more than **3 pages** in double columns (refer to IEEE format) excluding reference lists. The report could be simply titled as "lab report" but the student could also title the report in their own favourite way. The report does not need a cover page, the student should write their names under the title with student ID provided. The student should also name their assigned TA under their own names. Though literatures could be cited to support the ideas in the report, it is NOT necessary to review relevant literatures in the report hence it is absolutely no problems to cite no papers in the report. The lab report should be handed in with a single PDF file without source code attached. The use of Latex is strongly recommended.

The use of ChatGPT is allowed for proofreading and brainstorming ONLY. However, copying an AI generated solution to the task will not guarantee that you could pass the coursework. It is the critical thinking, experiment design and the analysis of result that matter. You MUST fully understand your code and the experiment you have designed in this coursework.

Marking Criteria

Lab report:

Editorial & Language Issues (10 marks)

10 marks: no formatting problems found.
8 marks: minor language issues or minor formatting problems.
6 marks: the report is generally fine with a few language issues and formatting problems.
4 marks: the report is barely readable.
2 marks: the report is hard to read but understandable.
0 marks: the report is not understandable.

Task 1, 2 and 3 (60 marks in total, each task worth 15 marks)

15 marks: scientific hypothesis has been demonstrated via the results presented.
12 marks: the results of different experiment configurations are compared and analysed.
9 marks: the results of different experiment configurations are compared with depth.
6 marks: the target of the task has been fully fulfilled.
3 marks: the target of the task has not been partially fulfilled with good attempts.
0 marks: the target of the task has not been fully fulfilled with no sensible attempts made.

Live Demonstration:

Answer Questions (15 marks in total, each question worth 5 marks)

5 marks: show full understanding of concepts with satisfactory answers provided.
4 marks: satisfactory answers provided.
3 marks: satisfactory answers provided with minor misunderstanding.
2 marks: the answer is barely satisfactory.
1 mark: the answer is not correct.
0 marks: the student is unable to answer the question.

Code Running (15 marks)

15 marks: may implement the code in an efficient way and predict the result with good understanding of algorithms.
12 marks: may implement the code as required and discuss the result in depth.

9 marks: may implement the code with assistance and show understanding of the result.
6 marks: may implement the code with assistance in a period and show some understanding of the result.
3 marks: cannot implement the required changes with a sensible expectation of result.
0 marks: cannot understand the intention of the required changes.

N.B. For cases that are not listed here, TA would match the criteria in the list and give a mark to the code running session. TA also have the right to mark suspected plagiarism and raise the case to module leader.

Award Marks (capped at 100 marks in total, no individual award cap)

+10 marks: demonstrate a novel scientific hypothesis in any tasks.
+5 marks: present the experiment in a way that the experiment could be re-implement by others with ease.
+5 marks: the report is formatted in a publishable way.

Penalties:

-10 marks: for improper citation
-20 marks: for severe improper citation (multiple improper citation or duplication of a whole paragraph)
- University academic integrity penalties apply.

Submission

1. Only submissions in PDF format are accepted.
2. Submit your lab report via the dedicated Learning Mall coursework link before the Friday of week 11.
3. Please name your submission file as ID_FirstName_LastName_C2.pdf (e.g., 1234567_FirstName_Surname_C2.pdf).
4. Late submission policy of XJTLU applies.