High Distinction TASK

About this task

Step-1

This task is designed to assess the High Distinction level expectations. There are two sets of evidence requested in this task: one is for SIT307 students and the other one for SIT720 students. Please select the set based on your unit code. DO NOT SUBMIT BOTH SETS OF EVIDENCE.

Step-2

Your tutor will then review your submission and will give you feedback. If your submission is incomplete the tutor will ask you to include missing parts. Tutor can also ask follow-up questions, either to clarify something that you have submitted or to assess your understanding of certain topics.

Feedback and submission deadlines

Feedback deadline: Friday 2 June (No submission before this date means no feedback!)
Submission deadline: Before creating and submitting portfolio.

Aim: The goal for this HD task is to reproduce the main results of a research paper "Mushroom data creation, curation and simulation to support classification tasks" or/and propose your own approach to do further improvement. You should read the paper carefully and get familiar with the mushroom dataset before doing the following tasks.

Evidence of Learning – SIT307

Execute your code into a jupyter notebook (.ipynb file) and keep the output, write a report (.pdf file) to answer the following questions, and submit your code and report to OnTrack.

Read the article and reproduce the results presented in Figure 4 using Python modules and packages (including your own script or customised codes). Write a report summarising the dataset, used ML methods, experiment protocol and results including variations, if any. During reproducing the results:

i) you should use the same set of features used by the authors.

ii) you should use the same classifier with exact parameter values.

iii) you should use the same training/test splitting approach as used by the authors.

iv) you should use the same pre/post processing, if any, used by the authors.

v) you should report the same performance metric (Accuracy and F2 score) as shown in Figure 4.

Evidence of Learning – SIT720

Execute your code into a jupyter notebook (.ipynb file) and keep the output, write a report (.pdf file) to answer the following questions, and submit your code and report to OnTrack.

1. Read the article and reproduce the results presented in Figure 4 and Figure 5 using Python modules and packages (including your own script or customised codes). Write a report summarising the dataset, used ML methods, experiment protocol and results including variations, if any. During reproducing the results:

i) you should use the same set of features used by the authors.

ii) you should use the same classifier with exact parameter values.

iii) you should use the same training/test splitting approach as used by the authors.

iv) you should use the same pre/post processing, if any, used by the authors.

v) you should report the same performance metric (Accuracy, F2 score, ROC) as shown in Figure 4 and Figure 5.

2. Design and develop your own ML solution for this problem. The proposed solution should be different from all approaches mentioned in the provided article. This does not mean that you must have to choose a new ML algorithm. You can develop a novel solution by changing the feature selection approach or parameter optimisations process of used ML methods or using different ML methods or adding regularization or different combinations of them. This means, the proposed system should be substantially different from the methods presented in the article but not limited to only change of ML methods. Compare the accuracy/F2 result with reported methods in the article. Write in your report

summarising your solution design and outcomes. The report should include:

i) Motivation behind the proposed solution.
ii) How the proposed solution is different from existing ones.
iii) Detail description of the model including all parameters so that any reader can implement your model.
iv) Description of experimental protocol.
v) Evaluation metrics.
vi) Present results using tables and graphs.
vii) Compare and discuss results with respect to existing literatures.
viii) Appropriate references (IEEE numbered).