HD Task

This task is designed to assess the High Distinction level expectations.

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: Not applicable

Submission deadline: Before creating and submitting portfolio.

Required documents

1. Submit a report (pdf format) in Ontrack (https://ontrack.deakin.edu.au)
2. Complete the problem credit task and submit your code file (.ipynb) separately in the OnTrack (https://ontrack.deakin.edu.au).

Background

The global health crisis of undiagnosed diabetes demands urgent attention. This chronic metabolic disorder, affecting millions worldwide, silently progresses without proper diagnosis. Early detection is crucial to prevent or delay serious complications like heart disease, blindness, and kidney failure. By raising awareness about the risk factors and symptoms of diabetes, and by promoting accessible screening programs, we can significantly improve early diagnosis rates and empower individuals to manage their health effectively.

Datasets Description

Download the PIMA Indians diabetes dataset from the given articles
(https://www.sciencedirect.com/science/article/pii/S240584402400567X).

Evidence of Learning- SIT307

1. Read the article and report the results (Accuracy, Precision, Recall, F-Measure) for PIMA Indians diabetes dataset using following classification methods:
   - Decision Tree
   - SVM
   - Random forest
   - Stacking Ensemble

Compare your result with **Table 4** (**Performances (%) of ML algorithm on pima dataset using train-test and CV**) of the manuscript and should be used for comparison purposes, if required. Write a report summarising the dataset, 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.

N.B.

(i) If you find any issue in reproducing results due to incomplete description of model in the provided article, then make your own assumption and explain the reason. If your justification is correct, then your solution will be considered correct and assessed accordingly.

(ii) Similarly, variation in results due to randomness of data splitting will also be considered during evaluation based on your explanation.

(iii) Obtained marks will be proportional to the number of ML methods that you will report in your submission with correctly reproduced results.

(iv) Make sure your submitted Python code segment generates the reported results, otherwise you will receive zero marks for this task.

2. Design and develop your own ML solution for this problem. The proposed solution should be different from all approaches mentioned in **Q-1**. 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 using different ML methods 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 result with reported methods in the article. Write a technical report summarising your solution design and outcomes. The report should include:
   - Detail description of the model including all parameters so that any reader can implement your model.
   - Description of experimental protocol.
   - Evaluation metrics.
   - Present results using tables and graphs.
   - Compare and discuss results with respect to existing literatures.

- Appropriate references (<u>IEEE numbered</u>).

3. Present your result in a 3 minutes video using PowerPoint slides/animation.