Distinction Task -I (Supervised learning)

Step-1

This task is designed to assess the Distinction level expectations.

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 20 September (No submission before this date means no feedback!)

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 growing trend of internet-connected appliances, or IoT devices, in our homes brings both convenience and security risks. These devices can be vulnerable to attacks that steal our data, disrupt their functionality, or even cause physical damage. To combat this, it's crucial to classify these attacks and implement effective defenses. Based on this scenario, the goal is to develop a prediction model that minimises a total attack attack related damage and save IoT enable devices from intruder.

Datasets Description

The attributes are as follows: target, then anonymised operational data. In total there are 84 attributes. Missing values are denoted by "na".

Datasets: “Dataset4.csv”

Evidence of Learning- SIT307

1. What are the differences between hyperparameter and parameter of a machine learning (ML) model. Explain your answer using at least two machine learning models that you have learned in this unit.
2. Prove that Elastic net can be used as either LASSO or Ridge regulariser.
3. Analyse the importance of the features for predicting “target'' using two different approaches. Explain the similarity/difference between outcomes.
4. Create three supervised machine learning (ML) models except any ensemble approach for predicting “target''.
    a. Report performance score using a suitable metric. Is it possible that the presented result is an overfitted one? Justify.
    b. Justify different design decisions for each ML model used to answer this question.
    c. Have you optimised any hyper-parameters for each ML model? What are they? Why have you done that? Explain.
    d. Finally, make a recommendation based on the reported results and justify it.
5. Build three ensemble models for predicting “target''.
    a. When do you want to use ensemble models over other ML models? Explain based on the models that you have used in Q4 and Q5.
    b. What are the similarities or differences among used ensemble models used in Q5?
    c. Is there any preferable scenario for using any specific model among the set of ensemble models?
    d. Write a report comparing performances of models built in question 4 and 5. Report the best method based on model complexity and performance.
    e. Is it possible to build ensemble model using ML classifiers other than decision tree? If yes, then explain with an example.