Pass Task 3

Solve the following set of problems using Python and submit the code file with the extension .ipynb in Olympus as part of your pass activity.

1. Load "digits" datasets from SKlearn. Classify digit classes using KNN. Use the same data splitting and performance metrics that you have used in previous week (week 4). Report your findings including comparison of results with week 4.

2. Create digits classification model using DT algorithm using 50-50% and 70-30% data splitting methods. Compare performances of these two models and explain the impact of difference in data splitting on the performances of the model.

3. Create two more KNN-based classification models using the dataset used in Q1 by varying distance metrics such as using **cityblock** and **cosine**. Report the performances of the developed models including Q1 and explain the similarity or differences if any.

4. Creating random forest model using HR-Employee-Attrition.csv dataset and improve the result using hyperparameter tuning. Hints. Visualise your performance fluctuation for different hyperparameter values.

5. Creating GradientBoost model using HR-Employee-Attrition.csv dataset and improve the result using hyperparameter tuning. Hints.

6. Compare the best model after hyperparameter tuning found in Q4 and Q5, and explain which model is good and why.

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.