PASS TASK (SVM models and KNN)

About this task

Step-1
At the completion of week 7 and 8 modules, you are required to complete a lesson review to tell us what you learnt and how you learnt it by submitting evidence requested at the end of this file.

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

Submission deadline: Before creating and submitting portfolio.

Evidence of Learning

1. **Submit a report (pdf format) in Ontrack (https://ontrack.deakin.edu.au)**
   1. Summarise the main points that is covered in weeks 7 and 8.
   2. Provide summary of your reading list – external resources, websites, book chapters, code libraries, etc.
   3. Reflect on the knowledge that you have gained by reading contents of this week with respect to machine learning.
   4. Attempt the quiz given in weekly content (7.11 and 8.14) and add screenshot of your score (≥85% is considered completion of the task) in this report.
2. **Complete the problem solving task and submit your code file (.ipynb) separately in the OnTrack (https://ontrack.deakin.edu.au).**

Evidence of Learning- SIT307

1. Read the "Dataset3.csv" and print the dimension of the dataset. Apply PCA on the dataset and select first three components except "Target". Print the dimension of modified dataset and visualise the data using appropriate plotting tool/s.

2. Create at least three SVM models to classify "graduation", "dropout", and "enrolled" students based on given information. Make your own choice for train-test data splitting and performance metrics. Use the same train-test split and performance metrics for all models. Report the performances and the used model hyper-parameters. Explain the reason behind your selection of data splitting approach and performance metrics.

3. Based on the model hyper-parameters used in Q2 share your understanding of hyper-parameters tuning in ML model development.

4. Repeat the same classification task done in Q2 using KNN (build two models). Use the same data splitting and performance metrics that you have used in Q-2. Report the performance and compare results with the findings of Q2.

5. Create two more KNN-based classification models by varying distance metrics. Report the impact of distance metrics on performances KNN models. Explain the necessity of tuning distance metrics during KNN model development with respect to your findings.