PASS TASK (Task 2.1)

About this task

Step-1

At the completion of week 2 modules, you are required to compete a lesson review to indicate what you have 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 25 July (No submission before this date means no feedback!)
Submission deadline: Before creating and submitting portfolio.

Evidence of Learning

1. Submit a summary report (pdf format) in Ontrack (https://ontrack.deakin.edu.au)
   1.1. Summarise the main points that is covered in week 1 and 2.
   1.2. Provide summary of your reading list – external resources, websites, book chapters, code libraries, etc.
   1.3. Reflect on the knowledge that you have gained by reading contents of this week with respect to machine learning.
   1.4. Attempt the quiz given in weekly content (1.28 and 2.14) and add screenshot of your score (>85% is considered completion of this task) in this report.
2. Complete the problem solving task given below and submit your code file (.ipynb) separately to OnTrack (https://ontrack.deakin.edu.au).

Problem Solving Task

1. Read "Microclimate sensors data" and print the feature name with numbers of missing entries.
2. Fill in the missing entries. For filling any feature, you can use either the mean or median value of the feature values from observed entries. Explain the reason behind your choice and print replacement value of each feature.
3. Use a histogram to show the distribution of the variable “PM25” and “PM10”. Explain the two distributions and show if there is a correlation between them.
4. Split the variable “LatLong" using an appropriate encoding approach and display the encoded values. Justify your selection of encoding approach.
5. From question 4, Apply the min-max scaling on continuous features. Plot distribution of these features before and after scaling. Is there any difference? Please explain.