

PASS TASK (Introduction to ML and Data wrangling)

About this task

Step-1

At the completion of week 1 and 2 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 26 July (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 1 and 2.
 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 quizzes given in weekly content (**1.28 and 2.14**) and add screenshot of your score ($\geq 85\%$ 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. Create three variables named 'Database name', 'Company name', and 'Company registration Number'. Add 5 random values in these variables, where added values are not replaceable and print them.
2. Create a dictionary variable with three keys, which are named same as the three variables created in the previous problem. Copy the variables' values of Q1 in the newly defined dictionary.
3. Define a string variable 'species' and assign value either 'Mammals' or 'Birds'. Print "Bald Eagle" if the value of species variable is 'Birds' otherwise print "Kangaroo". What will your program print if species='Reptiles'?
4. Is it possible to break a loop before executing for the defined number of iterations? If yes, provide an example using While loop structure. Otherwise, explain your answer (why and how).
5. Write a code snippet to print odd and even positioned characters of your name in two separate lines.
6. Import the "Dataset1.csv " file as a matrix. The row of the matrix represents data point or instance, and the column represents the variable or feature. Print the number of data instances and variables present in this dataset. Calculate mean of the third variable and deduct this mean from each instance of the third variable. Print both original and modified values of the variable.
7. Read "Dataset1.csv " and 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.
 - I. Encode the categorical variable "class" using an appropriate encoding approach and display the encoded values. Justify your selection of encoding approach. Save the modified data into a new csv file. Upload this new file with your submission in OnTrack.
 - II. Apply the min-max scaling on features. Plot distribution of first six features before and after scaling. Is there any difference? Please explain.