IFN646 - Portfolio item 5

Key points

- Individual assignment
- 6 marks
- Release date: 12 October 2023
- Due date: 29 October 2023

Overview

For this portfolio item, you will continue to work on microbial abundance, but your focus will switch to classification. Using the same datasets as in portfolio item 4, you will develop some code, and upload your analysis to Canvas.

We strongly recommend using a notebook and submitting an HTML or PDF export. This could be from a Jupyter environment on your own machine, a Google Colab notebook, etc.

1. Data for the portfolio item

You are working with the same three files as before:

- A BIOM table that contains abundance counts for various OTUs across a number of samples from the throat and nose of smokers and non-smokers.
- A metadata file that provides additional information about each sample.
- A taxonomy file that allows you to link the ID of each OTU to an actual name.

If you no longer have the data, it can be downloaded again from Canvas, in the folder for portfolio item 4.

2. Tasks

Overall goal

The overall goal of the analysis is to build simple classifiers to predict whether a given person is a smoker.

You are free to use any machine learning approach that you deem appropriate for this goal. The goal of the portfolio is not to develop very complex approaches. It has about identifying suitable, simple approaches and using them appropriately.

Task 1 [1.5 mark]

Detail and justify your choice of approach. [1 mark]

Clearly explain your strategy for splitting your data into training and testing test. [0.5 mark]

Task 2 [1.5 marks]

Build a classifier that predicts whether a person is a smoker *based on only nose samples*, show your results, and discuss the performance of the classifier.

Task 3 [1.5 marks]

Build a classifier that predicts whether a person is a smoker *based on only throat samples*, show your results, and discuss the performance of the classifier.

Task 4 [1.5 marks]

Build a classifier that predicts whether a person is a smoker *based on the two sample types*. This could be an approach that combines the classifiers from the previous two tasks, or a new classifier. Show your results, and discuss the performance of the classifier.

3. Submission

You will submit your analysis through Canvas using the "Portfolio Item 5" link.

We strongly recommend using a notebook, exporting your work as an HTML and PDF file, and uploading this as your submission.

Alternatively, if you prefer not to use Jupyter or Colab, you can submit a PDF report that contains both your answers for each task and your Python code as appendix.

Submission will close at 11.59pm on the due date.

4. Academic honesty

This is an individual assessment, and you need to submit your own work. Standard plagiarism checks will be performed on your code. We also reserve the right to select some submissions and ask students to explain the reasoning behind their answers.