Project Stage 2 FAQs

What if I my proposed approach doesn't produce a positive result?

A negative result is fine and won't affect your mark. The assignment should demonstrate an understanding of framing and interpreting experiments. It is as important to know when something doesn't work. And it is critical to analyse why the setup or approach didn't work out.

How can I test the reliability of a cluster analysis?

Good question. It doesn't make sense to do post-hoc testing of the results of an exploratory data analysis. However, we can assess reliability in other ways, e.g., test for evidence of clustering, perform a bootstrap stability analysis, use cophenetic correlation for hierarchical clustering.

How can I test the reliability of a single classifier?

One option is to calculate confidence intervals around the mean using, e.g., the bootstrap procedure covered in week 7 lecture and exercise. Alternatively, scikit-learn implements regression for prediction, for which we can calculate r2 and prediction intervals. It is also possible to derive p-values or use other statistical packages.

Does significance testing apply to classifiers?

Absolutely. There are two common research questions: Does my new model outperform the baseline? Which of several models performs best? Both can be tested using pairwise tests covered in week 7 lecture and exercise, e.g., paired Student's t-test, Wilcoxon signed-ranks test.

Do I need to do a live demonstration?

No. The final presentation should be an oral presentation with slides. In addition to being part of your mark, it's an opportunity to share the outcome of your project with the class.