Case II Description

!NOTE: The data shared in this case are governed by licences. Please refrain from uploading the data to an LLM or other cloud based services, also those that are hosted in Denmark and those that DTU has licenses to!

The case *consists* of doing an exploratory (unsupervised) analysis using the data source or parts of the data source, **but is not limited to only unsupervised-decomposition methods**. You are free to explore any one of the tools covered between weeks 8-12. The problem formulation here is more open-ended than case I, to allow you the choice to pursue a direction of investigation. Furthermore, real-world applications often require working with non-standardized data, and designing an ML-pipeline from data curation and pre-processing to evaluation is part of the solution, as much as model development. In other words, points in the grading will also be given for the overall design-choices of the ML-pipeline. You are free to choose **one** of the following three topics and continue your exploration along that for the case:

- Explore clustering methods to derive insights from the data.

- Use decomposition and subspace methods to investigate low-dimension representation of the data.

- Employ Canonical correlation analysis or other supervised methods to model, for eg: the response (emotions) using the features as input.

If you would like pursue to investigation using methods besides the above, feel free to do so after running a sanity check with the instructor. You will be using the EmoPairCompete dataset which is described in *case_description_2025.pdf*. Features have been extracted from the dataset and shared in the HR_data.csv or the updated on HR_data_2.csv; While you are *expected* to use the features in this file as the absolute minimum, You can also use the time-seies data under dataset folder if that fits your modelling goals. Note that the raw BVP signal is noisy, which is why no features from this signal are provided in the HR_data file.

An interpretation of the model should be present. You are allowed to use other data sources as well. You may supplement your unsupervised analysis with a supervised analysis, if you wish. Formulate a clear purpose for your project, so the aim of your analysis is clear. What are your research questions?

What kind of representation would you like to derive from the features of the biosignals? What could it be used for? We expect you to use methods from the course curriculum. Please make solid arguments for your choice of methods. The choice of methods should align with your aim. The evaluation criteria for the case is provided in *case_evaluation_2025.pdf*