

User Manual

The web frontend is designed to provide an intuitive exploration of the different datasets, clustering algorithms and distances. In an interactive interface multiple clustering settings can be chosen and a visualisation of the results is directly generated. A cluster-table is used to store previous calculated results, which can be plotted within the evaluation module.

A checkbox at the top of the page (see Figure 1 A), which is set by default, allows the use of precalculated clustering results, which have been computed beforehand (with a random seed value) and are stored in the github repository. This was done for reproduction purposes. The second checkbox (see Figure 1 B) allows the option for interactive projection plots.

The user can choose between four datasets (see Figure 1 C), four distance measures (see Figure 1 E) and four different algorithms (see Figure 1 D). If kmeans, kmedian or kmedoids is chosen a value for the parameter k between 1 and 10 has to be set (see Figure 1 F) The default value for k is 3. If DBSCAN is chosen the user has to define a value for epsilon and the minimal number of nearest points. The parameter settings can be adjusted with an interactive slider widget.

Datascience: Group 42

- A** ☒ Use precalculated results (with random seed for reproduction).
- B** ☒ Use interactive altair plots over static seaborn plots.

Settings

C	Choose a beautiful dataset iris	Choose a lovely clustering algorithm kmeans	D
E	Choose an awesome distance measure euclidean	Choose a nice value for k (number of clusters) 2 3 10	F

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Figure 1: First part of the web frontend. Setting options for clustering parameters.

Moreover the perplexity value for the t-SNE projection can be set individually between 5 and 50 with a slider (see Figure 2 A). The default value is 25. The t-SNE and PCA plots are shown next to each other to allow a direct comparison of the lower-dimensional projections. To virtually interact with the plots, the button shown in Figure 2 C can be clicked. Please note that this option is only available when the checkmark shown in Figure 1 B is set. Furthermore, this button provides the ability to save the plot in different formats. The selected dataset can be viewed in a table format below the plots (see Figure 2 B). The frontend is reloaded entirely if a parameter value is changed, a different setting is made or a button is clicked.

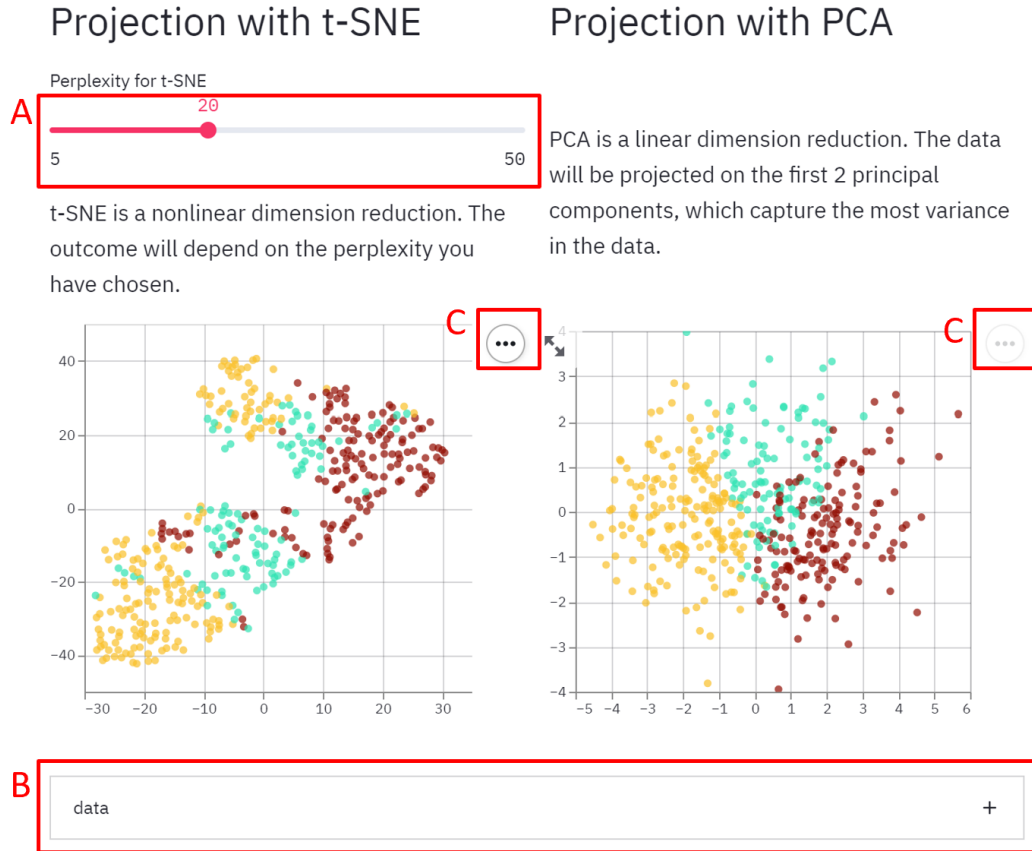


Figure 2: Second part of the web frontend. Projection results of a clustering.

For the evaluation of the clustering results, a clustering index can be chosen as shown in Figure 3 C. Afterwards, the clustering result can be saved through addition to the clustering table with the *add*-button (see Figure 3 A) and the selected index will be calculated. To compare this index to the resulting indices of clusterings with other settings, the *add*-button can be clicked repeatedly after desired adjustment of the clustering settings. To start over and compare further clustering indices, the *reset*-button (see Figure 3 B) can be clicked. The previous display of resulting indices will be cleared as well as the clustering table.

Clustering evaluation

Clustering results can be stored in a cluster-table and used for comparative evaluation.

A **B**

Cluster-table cleared succesfully!

Cluster-table is empty!

Choose an adorable index

C

Plot not possible.

Figure 3: Third part of the web frontend. Evaluation of clustering results.

Ancillary streamlit settings can be found at the upper right corner. Additional explanatory texts provide help and more information.