Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

Load the data from input file: Create set of candidate models (for set range of clusters number); Choose initial centroids: for each model: random/k++: Select one of them based on Dunn/Silhouette: Choose distance metrics for the given dimensionality/user's wish (Euclidean, correlation or city-block); Skip less important dimensions (PCA); Save the result into the output file and visualise it

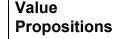
ML task



Input, output to predict. type of problem.

Input: data of samples x features character Output: array of labels

The problem is grouping one.





What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

We want to give end-user a way to find patterns in data.

Data Sources



Which raw data sources can we use (internal and external)?

Text file given by user.



Collecting Data



How do we get new data to learn from (inputs and outputs)?

No new data from the user. grouping is based on labels of other samples

Making **Predictions**

(optional).

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

The moment we load the file with data we make predictions for it. Depending on the size of data.

Offline **Evaluation**

Methods and metrics to evaluate the system before deployment.

Inertia, number of iterations. Silhouette coefficient/Dunn index, visualisation (optional)

Features

Input representations extracted from raw data sources.

The data given in the columns of input table; The reduced form of data by **PCA** for visualisation (optional)

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

One model per data, at the moment of runtime. The estimator can be reused for a different data depending on the needs of the user.

Live Evaluation and Monitorina

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

None



