



**Robust,
useful
pipeline**

**Export /
download**

Inputs: Parameters in the pipeline defining what needs to be downloaded.
Outputs: Raw data, saved in an easily accessible location (e.g. locally)

Engineering

Inputs: Raw data exported above
Outputs: Arrays, split into test and train sets, which can be directly fed into a machine learning model

Model Training

This step may manipulate the data further (depending on the model being used), but should result in predictions for the test set and a saved, trained model.

Analysis

Take the trained model and the predictions, and explain them. In addition, there should be tools here to explore how good the model's predictions are.

#ESoWC2019