

Scikit-learn LEVEL 8

Working With Text Data

Tutorial setup

Loading the 20 newsgroups dataset

Extracting features from text files

Tokenizing text with scikit-learn

From occurrences to frequencies

Training a classifier

Building a pipeline