

PREPROCESSING

pdf2nlp.py

Text Mining
of pdf articles

NLP Pipeline

TRAINING

mlpipe.py

Select features

Extract features
and hash to term-
document matrix

Train machine
learning algorithm

PREDICTION

loadmodel.py

Extract features
and hash to term-
document matrix

Machine learn-
ing model

LABEL

