

# Email classification

Tuesday, October 30, 2018 1:16 PM

<https://github.com/wroberts/pyqermanet>

<https://datascience.blog.wzb.eu/2017/05/19/lemmatization-of-german-language-text/>

I have not checked syn\_dict as per original code

<https://github.com/WZBSocialScienceCenter/germalemma>

<https://github.com/Bachfischer/german2vec>

<https://textminingonline.com/tag/german-word2vec>

<https://github.com/sidsriv/Applied-text-mining-in-Python/blob/master/Assignment%2B1.ipynb>

Date parser

<https://dateparser.readthedocs.io/en/latest/>

<https://dateparser.readthedocs.io/en/latest/>

pipeline

<https://towardsdatascience.com/machine-learning-for-text-classification-using-spacy-in-python-b276b4051a49>

<https://www.kaggle.com/baghern/a-deep-dive-into-sklearn-pipelines>

<https://www.kaggle.com/gautham11/building-a-scikit-learn-classification-pipeline>

<https://www.kaggle.com/metadist/work-like-a-pro-with-pipelines-and-feature-unions>

<https://iaml.it/blog/optimizing-sklearn-pipelines>

<https://stackoverflow.com/questions/42266737/parallel-pipeline-to-get-best-model-using-gridsearch>

<https://ramhiser.com/post/2018-03-25-feature-selection-with-scikit-learn-pipeline/>

##function creation for pipeline

<https://ryan-cranfill.github.io/sentiment-pipeline-sklearn-3/>

<https://towardsdatascience.com/multi-label-text-classification-with-scikit-learn-30714b7819c5>

RF

<https://datascience.stackexchange.com/questions/6683/feature-selection-using-feature-importances-in-random-forests-with-scikit-learn>

[https://scikit-learn.org/stable/auto\\_examples/compose/plot\\_column\\_transformer.html](https://scikit-learn.org/stable/auto_examples/compose/plot_column_transformer.html)

<https://scikit-learn.org/stable/modules/compose.html>

<https://towardsdatascience.com/how-i-used-machine-learning-to-classify-emails-and-turn-them-into-insights-efed37c1e66>

<https://www.kaggle.com/ratatman/analyzing-multilingual-data>

Feature reduction:-

[https://scikit-learn.org/0.18/auto\\_examples/plot\\_compare\\_reduction.html](https://scikit-learn.org/0.18/auto_examples/plot_compare_reduction.html)

`PCA(iterated_power=7), NMF(),  
SelectKBest(chi2)`

From [<https://scikit-learn.org/0.18/auto\\_examples/plot\\_compare\\_reduction.html>](https://scikit-learn.org/0.18/auto_examples/plot_compare_reduction.html)

From [<https://scikit-learn.org/0.18/auto\\_examples/plot\\_compare\\_reduction.html>](https://scikit-learn.org/0.18/auto_examples/plot_compare_reduction.html)