

Machine Learning Project - 2020

Group 04

Harry Chirayil, Christopher Keim, Stefan Schmutz

Report and self-contained MATLAB code with data

This folder contains everything which will be handed in for grading.

It consists of the written report (`report.pdf`) and the code and data used to get from data to result (`report_code_full.mlx`).

Since this version of the script is time consuming and fails when not enough RAM is available, there's another light-weight version (`report_code_efficient.mlx`) which executes the same code with a reduced dataset to decrease runtime and computational requirements.

Document structure

```
.
+-- README.pdf
+-- data
|   +-- train_features.csv
|   +-- train_labels.csv
|   +-- test_features.csv
|   +-- test_labels.csv
|   +-- vocabulary.csv
|   +-- vocabulary_stop_words.csv
+-- document_term_matrix.m
+-- cross_validation.m
+-- report_code_efficient.mlx
+-- report_code_full.mlx
+-- report.pdf
```

Execute MATLAB code

Make sure the current folder is set to the one this README.pdf document is located and the workspace is cleared (`clear all; close all; clc`).

Execute codeblocks of `report_code_full.mlx` or if time and/or memory is restricted `report_code_efficient.mlx` . The required functions are stored in `document_term_matrix.m` and `cross_validation.m` which are called from the `.mlx` scripts and do not need to be executed separately.