SEMINAR
KLASSIFIKATION & CLUSTERING

ERGEBNISSE BASELINE

Stefan Langer

stefan.langer@cis.uni-muenchen.de

Letter – author – logistic regression

#Performance:
Seconds used for training: 1156
Seconds used for classification: 8

#Classification report:

	precision	recall	f1-score	support
Franz Kafka	0.907	0.911	0.909	280
Friedrich Schiller	0.774	0.861	0.815	266
Henrik Ibsen	0.999	0.977	0.988	897
James Joyce	0.952	0.911	0.931	682
Johann Wolfgang von Goethe	0.811	0.658	0.726	228
Virginia Woolf	0.968	0.986	0.977	1901
Wilhelm Busch	0.925	0.965	0.945	627
accuracy			0.945	4881
macro avg	0.905	0.895	0.899	4881
weighted avg	0.945	0.945	0.944	4881

Letter – author – pytorch simple

```
#Performance:
Seconds used for training: 550
Seconds used for classification: 1

#Classification report:
                          precision    recall  f1-score   support

              Franz Kafka      0.932     0.932     0.932       280
        Friedrich Schiller     0.706     0.910     0.795       266
              Henrik Ibsen     0.989     0.990     0.989       897
              James Joyce      0.922     0.947     0.934       682
Johann Wolfgang von Goethe     0.830     0.601     0.697       228
            Virginia Woolf     0.983     0.972     0.977      1901
             Wilhelm Busch     0.974     0.955     0.965       627

                  accuracy                         0.947      4881
                 macro avg     0.905     0.901     0.899      4881
              weighted avg     0.949     0.947     0.946      4881
```

Letters – language – logistic regression

```
#Performance:
Seconds used for training: 848
Seconds used for classification: 8

#Classification report:
              precision    recall  f1-score   support

          da      0.995     1.000     0.998       844
          de      0.999     0.997     0.998      1448
          en      0.998     0.999     0.998      2519
          fr      0.973     0.947     0.960        38
          it      0.964     0.844     0.900        32

    accuracy                          0.997      4881
   macro avg      0.986     0.958     0.971      4881
weighted avg      0.997     0.997     0.997      4881
```

Letters – language – pytorch simple

```
#Performance:
Seconds used for training: 536
Seconds used for classification: 1

#Classification report:
              precision    recall  f1-score   support

          da       0.998     1.000     0.999       844
          de       1.000     0.998     0.999      1448
          en       0.999     1.000     0.999      2519
          fr       0.974     1.000     0.987        38
          it       0.967     0.906     0.935        32

    accuracy                          0.999      4881
   macro avg       0.987     0.981     0.984      4881
weighted avg       0.999     0.999     0.999      4881
```

News – category – logistic regression

```
#Performance:
Seconds used for training: 4577
Seconds used for classification: 33

#Classification report:
                precision    recall  f1-score   support

ARTS & CULTURE      0.565     0.309     0.399       282
  BLACK VOICES      0.505     0.429     0.463       392
      BUSINESS      0.466     0.416     0.439       380
       COLLEGE      0.686     0.505     0.582       212
             …
        TRAVEL      0.634     0.689     0.660       405
      WEDDINGS      0.779     0.775     0.777       400
    WEIRD NEWS      0.377     0.424     0.399       408
      WELLNESS      0.362     0.445     0.399       407
         WOMEN      0.433     0.448     0.440       400
         WORLD      0.612     0.656     0.633       404

      accuracy                          0.559     12824
     macro avg      0.565     0.551     0.554     12824
  weighted avg      0.566     0.559     0.559     12824
```

News – category – pytorch simple

```
#Performance:
Seconds used for training: 348
Seconds used for classification: 2

#Classification report:
                precision    recall  f1-score   support

ARTS & CULTURE      0.452     0.348     0.393       282
  BLACK VOICES      0.296     0.464     0.362       392
      BUSINESS      0.400     0.379     0.389       380
       COLLEGE      0.647     0.467     0.542       212
             …
        TRAVEL      0.547     0.667     0.601       405
     WEDDINGS       0.701     0.772     0.735       400
    WEIRD NEWS      0.397     0.287     0.333       408
      WELLNESS      0.287     0.568     0.382       407
         WOMEN      0.384     0.362     0.373       400
         WORLD      0.530     0.584     0.556       404

      accuracy                          0.507     12824
     macro avg      0.519     0.502     0.502     12824
  weighted avg      0.521     0.507     0.506     12824
```

Movie – sentiment – logistic regression

```
#Performance:
Seconds used for training: 406
Seconds used for classification: 26

#Classification report:
              precision    recall  f1-score   support

    negative      0.910     0.890     0.900      4985
    positive      0.893     0.913     0.903      5015

    accuracy                        0.902     10000
   macro avg      0.902     0.902     0.902     10000
weighted avg      0.902     0.902     0.902     10000
```

Movie – sentiment – pytorch simple

```
#Performance:
Seconds used for training: 581
Seconds used for classification: 3

#Classification report:
              precision    recall  f1-score   support

    negative      0.935     0.823     0.875      4985
    positive      0.843     0.943     0.890      5015

    accuracy                        0.883     10000
   macro avg      0.889     0.883     0.883     10000
weighted avg      0.889     0.883     0.883     10000
```