|  |  |
| --- | --- |
|  | Example of a text with different preprocessing |
| No SW and No ST | a revolução democrática atravessa um momento que pode ser decisivo os debates divergências |
| SW and No ST | revolução democrática atravessa momento pode ser decisivo debates divergências |
| No SW and ST | a revoluç democrá atravess um moment que pod ser decis os debat diverg |
| SW and ST | revoluç democrá atravess moment pod ser decis debat diverg |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Vocabulary Size | No SW and No ST | SW and No ST | No SW and ST | SW and ST |
| 1 gram | 27042 | 26935 | 11556 | 11524 |
| 1 + 2 gram | 339805 | 389755 | 254599 | 308794 |

|  |  |
| --- | --- |
|  | Class 0 |
| Precision | 1 |
| Model | FalseTrueFalse0.0010.52SVC.csv |
| Recall | 1 |
| Model | FalseFalseFalse0.00010.53BernoulliNB.csv |
| Support | 1380 |
| Model | FalseFalseFalse0.0050.51BernoulliNB.csv |
| F1\_Score | 0.77 |
| Model | TrueFalseFalse0.00010.992MultinomialNB.csv |

|  |  |
| --- | --- |
|  | Class 1 |
| Precision | 1 |
| Model | FalseFalseFalse0.0050.993GaussianNB.csv |
| Recall | 0.99 |
| Model | FalseFalseTrue0.00010.993SVC.csv |
| Support | 1003 |
| Model | TrueFalseTrue0.0010.51BernoulliNB.csv |
| F1\_Score | 0.57 |
| Model | TrueFalseFalse0.00010.993MultinomialNB.csv |

|  |  |
| --- | --- |
|  | Class 2 |
| Precision | 1 |
| Model | FalseFalseTrue0.00010.53SVC.csv |
| Recall | 0.89 |
| Model | FalseFalseTrue0.0050.992SVC.csv |
| Support | 801 |
| Model | FalseTrueTrue0.00010.993BernoulliNB.csv |
| F1\_Score | 0.55 |
| Model | TrueFalseFalse0.00010.992MultinomialNB.csv |

|  |  |
| --- | --- |
|  | Class 3 |
| Precision | 1 |
| Model | FalseFalseFalse0.00010.52BernoulliNB.csv |
| Recall | 0.88 |
| Model | FalseTrueTrue0.00010.53SVC.csv |
| Support | 382 |
| Model | FalseTrueTrue0.0050.53BernoulliNB.csv |
| F1\_Score | 0.56 |
| Model | TrueFalseTrue0.0010.992BernoulliNB.csv |

|  |  |
| --- | --- |
|  | Class 4 |
| Precision | 1 |
| Model | FalseTrueFalse0.00010.52BernoulliNB.csv |
| Recall | 0.89 |
| Model | FalseFalseFalse0.00010.991SVC.csv |
| Support | 279 |
| Model | TrueFalseFalse0.00010.991BernoulliNB.csv |
| F1\_Score | 0.48 |
| Model | TrueFalseFalse0.0010.52BernoulliNB.csv |

|  |  |
| --- | --- |
|  | Class 5 |
| Precision | 1 |
| Model | FalseFalseFalse0.00010.52MultinomialNB.csv |
| Recall | 0.87 |
| Model | TrueFalseFalse0.0050.52GaussianNB.csv |
| Support | 84 |
| Model | TrueFalseFalse0.0050.53BernoulliNB.csv |
| F1\_Score | 0.35 |
| Model | FalseFalseFalse0.0010.52MultinomialNB.csv |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | precision | recall | f1-score | support |
| 0 | .65 | .90 | .76 | 1347 |
| 1 | .63 | .53 | .57 | 929 |
| 2 | .66 | .53 | .59 | 756 |
| 3 | .74 | .56 | .64 | 350 |
| 4 | .70 | .38 | .49 | 226 |
| 5 | .94 | .23 | .37 | 69 |
| Avg/Total | .66 | .66 | .64 | 3677 |

precision recall f1-score support

0 0.65 0.90 0.76 1347

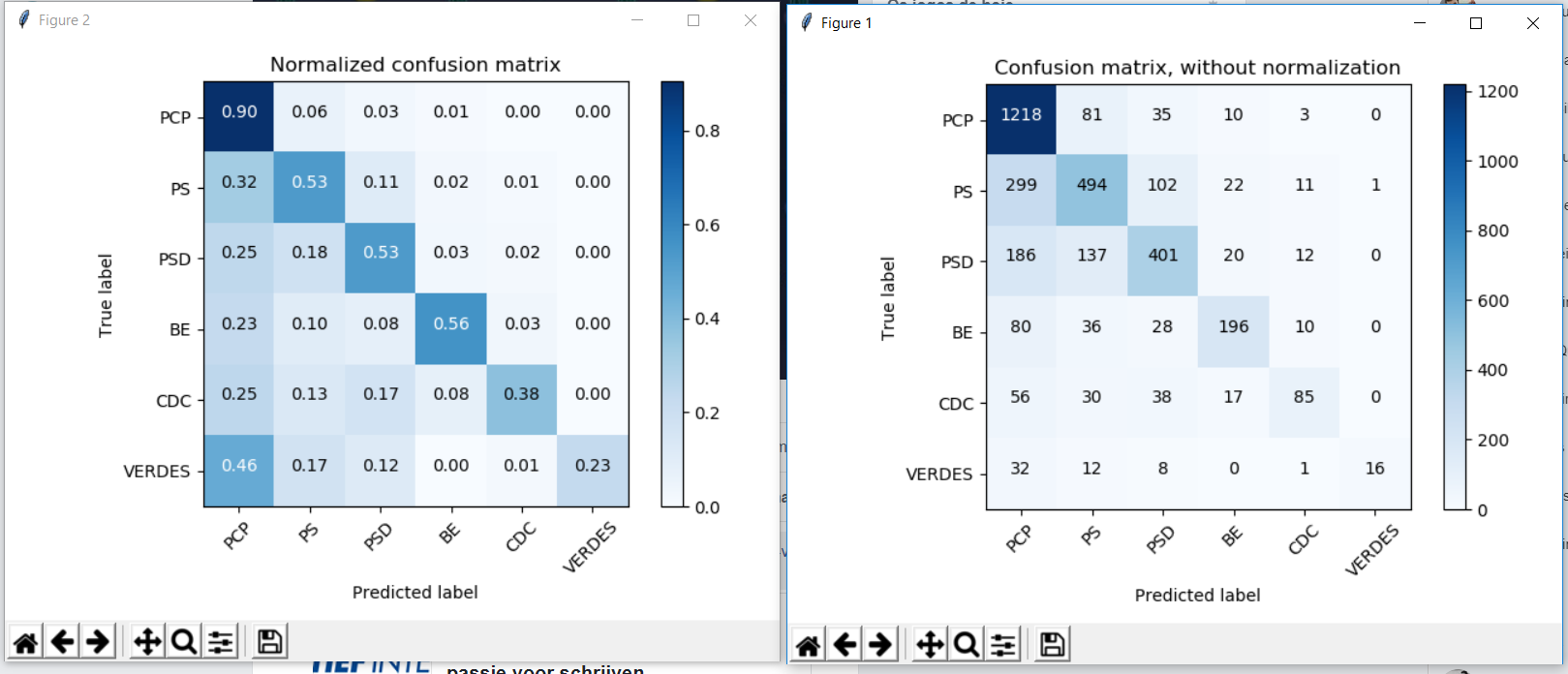
1 0.63 0.53 0.57 929

2 0.66 0.53 0.59 756

3 0.74 0.56 0.64 350

4 0.70 0.38 0.49 226

5 0.94 0.23 0.37 69

avg / total 0.66 0.66 0.64 3677 

Nao definir limite maximo levou a piores resultados.

Stoping words e Steeming tb nao ajudaram.

5 NGrams para cima tb nao ajudou

Emsemble methods decreased performance

