Results for both datasets with a Gaussian Naive Bayes (without the ‘Other’ class):

| South Park  Naive Bayes: Gaussian  Without Other | Baseline | Unigram | Bigram | Trigram | Improvement  of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,4 | 0,4 | **0,63** | **0,63** | 0,23 |
| Characters: 5 | 0,33 | 0,58 | **0,63** | **0,63** | 0,3 |
| Characters: 7 | 0,31 | 0,58 | **0,63** | **0,63** | 0,32 |
| Characters: 10 | 0,29 | 0,58 | **0,63** | **0,63** | 0,34 |

| Game of Thrones  Naive Bayes: Gaussian  Without Other | Baseline | Unigram | Bigram | Trigram | Improvement of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,45 | 0,49 | **0,65** | 0,57 | 0,3 |
| Characters: 5 | 0,3 | **0,65** | **0,65** | 0,57 | 0,35 |
| Characters: 7 | 0,24 | **0,65** | **0,65** | 0,57 | 0,41 |
| Characters: 10 | 0,2 | **0,65** | **0,65** | 0,57 | 0,45 |

After experimenting on a smaller part of the dataset, our findings are that the results when including the ‘Other’ class were not able to beat the baseline scores.

SVM

[:10000]

| South Park  Svm:  Without other | Baseline | Unigram | Bigram | Trigram | Improvement of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,4 | 0,53 | **0,67** | 0,61 | 0,27 |
| Characters: 5 | 0,33 | **0,71** | 0,67 | 0,61 | 0,38 |
| Characters: 7 | 0,31 | **0,71** | 0,67 | 0,61 | 0,4 |
| Characters: 10 | 0,29 | **0,71** | 0,67 | 0,61 | 0,42 |

| Game of Thrones  Svm:  Without other | Baseline | Unigram | Bigram | Trigram | Improvement of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,45 | 0,52 | **0,6** | 0,58 | 0,15 |
| Characters: 5 | 0,3 | **0,66** | 0,6 | 0,58 | 0,36 |
| Characters: 7 | 0,24 | **0,66** | 0,6 | 0,58 | 0,42 |
| Characters: 10 | 0,2 | **0,66** | 0,6 | 0,58 | 0,46 |

After experimenting on a smaller part of the dataset, our findings are that the results when including the ‘Other’ class were not able to beat the baseline scores.

KNN

[:10000]

| South Park  Knn:  Without other | Baseline | Unigram | Bigram | Trigram | Improvements of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,4 | 0,59 (K=19) | 0,61 (K=1) | **0,62 (K=1)** | 0,22 |
| Characters: 5 | 0,33 | 0,59 (K=19) | 0,61 (K=1) | **0,62 (K=1)** | 0,29 |
| Characters: 7 | 0,31 | 0,59 (K=19) | 0,61 (K=1) | **0,62 (K=1)** | 0,31 |
| Characters: 10 | 0,29 | 0,59 (K=19) | 0,61 (K=1) | **0,62 (K=1)** | 0,33 |

| South Park  Knn:  With other | Baseline | Unigram | Bigram | Trigram | Improvement of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,65 | **0,66 (K=37)** | **0,66 (K=19)** | 0,65 (K=21) | 0,01 |
| Characters: 5 | 0,58 | **0,79 (K=17)** | 0,78 (K=15) | 0,78 (K=3) | 0,21 |
| Characters: 7 | 0,55 | **0,79 (K=17)** | 0,78 (K=15) | 0,78 (K=3) | 0,24 |
| Characters: 10 | 0,52 | **0,79 (K=17)** | 0,78 (K=15) | 0,78 (K=3) | 0,27 |

| Game of Thrones  Knn:  Without other | Baseline | Unigram | Bigram | Trigram | Improvement of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,45 | **0,62 (K=61)** | 0,49 (K=25) | 0,6 (K=1) | 0,17 |
| Characters: 5 | 0,3 | **0,62 (K=61)** | 0,49 (K=25) | 0,6 (K=1) | 0,32 |
| Characters: 7 | 0,24 | **0,62 (K=61)** | 0,49 (K=25) | 0,6 (K=1) | 0,38 |
| Characters: 10 | 0,2 | **0,62 (K=61)** | 0,49 (K=25) | 0,6 (K=1) | 0,42 |

| Game of Thrones  Knn:  With other | Baseline | Unigram | Bigram | Trigram | Improvement of best model |
| --- | --- | --- | --- | --- | --- |
| Characters: 3 | 0,84 | **0,85 (K=9)** | **0,85 (K=7)** | **0,85 (K=7)** | 0,01 |
| Characters: 5 | 0,75 | **0,89 (K=9)** | **0,89 (K=5)** | **0,89 (K=7)** | 0,14 |
| Characters: 7 | 0,69 | **0,89 (K=9)** | **0,89 (K=5)** | **0,89 (K=7)** | 0,2 |
| Characters: 10 | 0,63 | **0,89 (K=9)** | **0,89 (K=5)** | **0,89 (K=7)** | 0,26 |