|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 0 | ~ | 93.07 | ~ | ~ |
| 1 | 52.06 | 91.29 | ~ | 60.43 |
| 5 | ~ | 66.19 | 93.55 | 51.99 |
| 10 | ~ | 59.95 | 94.31 | ~ |
| 20 | ~ | 57.06 | 94.99 | ~ |
| 50 | ~ | 52.67 | 94.65 | ~ |

**Accuracy % for Naïve Bayes with different thresholds for different feature sets (Train Data)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 0 | ~ | 63.55 | ~ | ~ |
| 1 | 52.68 | 63.14 | ~ | 53.73 |
| 5 | ~ | 59.86 | 60.38 | 52.68 |
| 10 | ~ | 57.00 | 61.74 | ~ |
| 20 | ~ | 57.00 | 62.16 | ~ |
| 50 | ~ | 52.96 | 62.51 | ~ |

**Accuracy % for Naïve Bayes with different thresholds for different feature sets (Test Data)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 0 | 0 | 4343 | 0 | 0 |
| 1 | 5 | 4058 | 0 | 280 |
| 5 | 0 | 339 | 4003 | 1 |
| 10 | 0 | 145 | 4198 | 0 |
| 20 | 0 | 68 | 4275 | 0 |
| 50 | 0 | 18 | 4325 | 0 |

**Number of features in different feature sets**

**Training and Test data are in the ratio 1:1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 0 | ~ | 92.96 | ~ | ~ |
| 1 | 60.06 | 91.69 | ~ | 66.58 |
| 5 | ~ | 68.77 | 93.35 | 60.06 |
| 10 | ~ | 63.25 | 94.36 | ~ |
| 20 | ~ | 60.80 | 94.09 | ~ |
| 50 | ~ | 59.93 | 93.74 | ~ |

**Accuracy % for Naïve Bayes with different thresholds for different feature sets (Train Data)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 0 | ~ | 94.32 | ~ | ~ |
| 1 | 52.68 | 93.04 | ~ | 76.01 |
| 5 | ~ | 75.09 | 94.69 | 72.34 |
| 10 | ~ | 73.99 | 95.05 | ~ |
| 20 | ~ | 71.79 | 95.05 | ~ |
| 50 | ~ | 71.79 | 94.69 | ~ |

**Accuracy % for Naïve Bayes with different thresholds for different feature sets (Test Data)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 0 | 0 | 5829 | 0 | 0 |
| 1 | 6 | 5421 | 0 | 402 |
| 5 | 0 | 495 | 5330 | 4 |
| 10 | 0 | 212 | 5617 | 0 |
| 20 | 0 | 97 | 5732 | 0 |
| 50 | 0 | 30 | 5799 | 0 |

**Number of features in different feature sets**

**Training and Test data are in the ratio 4:1**