|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Obscene junk words | Remaining words | Words below threshold | Words with Hashtags |
| 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 |
| 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 |
| 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**