|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Variations** | **N-gram Range** | **Dictionary Size** | **Test Accuracy** |
| *original (no extra processing of inputs)* | [1, 1] | 5000 | 0.9151357279956128 |
| *remove stop words* | [1, 1] | 5000 | 0.9145873320537428 |
| *remove punctuation* | [1, 1] | 5000 | 0.9158212229229503 |
| *stemming* | [1, 1] | 5000 | 0.9141760350973402 |
| *remove emojis* | [1, 1] | 5000 | 0.9152728269810804 |
| *convert to lowercase* | [1, 1] | 5000 | 0.9151357279956128 |
| *remove stop words, punctuation, and emojis, stemming, and convert to lowercase* | [1, 1] | 5000 | 0.9123937482862626 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Variations** | **N-gram Range** | **Dictionary Size** | **Test Accuracy** |
| *original (no extra processing of inputs)* | [1, 3] | 25000 | 0.9714834110227585 |
| *remove stop words* | [1, 3] | 25000 | 0.9317247052371812 |
| *remove punctuation* | [1, 3] | 25000 | 0.9713463120372909 |
| *stemming* | [1, 3] | 25000 | 0.9720318069646284 |
| *remove emojis* | [1, 3] | 25000 | 0.9716205100082259 |
| *convert to lowercase* | [1, 3] | 25000 | 0.9714834110227585 |
| *remove stop words, punctuation, and emojis, stemming, and convert to lowercase* | [1, 3] | 25000 | 0.9422813271181794 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Type** | **Feature Type** | **N-gram Range** | **Dictionary Size** | **Test Accuracy** | **Test BER** |
| *baseline* | positive negative set comparison | N/A | N/A | 0.6564299424184261 | 0.34777122618186473 |
| *logistic regression* | stemming | [1, 3] | 25000 | 0.9606525911708254 | 0.03956517119172476 |
| *SVC* | stemming | [1, 3] | 25000 | 0.972031806964628 | 0.028096151715940598 |
| *SVC* | stemming and newline character count | [1, 3] | 25000 | 0.9709350150808884 | 0.02923785351338759 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Type** | **Features Type** | **N-gram Range** | **Dictionary Size** | **Test Accuracy** |
| *SVC* | stemming | [1, 1] | 5000 | 0.9141760350973402 |
| *SVC* | stemming | [1, 1] | 25000 | 0.9159583219084179 |
| *SVC* | stemming | [1, 2] | 5000 | 0.9621606800109679 |
| *SVC* | stemming | [1, 2] | 25000 | 0.9707979160954209 |
| *SVC* | stemming | [1, 3] | 5000 | 0.9636687688511105 |
| *SVC* | stemming | [1, 3] | 25000 | 0.9720318069646284 |
| *SVC* | stemming | [2, 2] | 5000 | 0.9439265149437894 |
| *SVC* | stemming | [2, 2] | 25000 | 0.9642171647929806 |
| *SVC* | stemming | [2, 3] | 5000 | 0.9433781190019194 |
| *SVC* | stemming | [2, 3] | 25000 | 0.9628461749383055 |
| *SVC* | stemming | [3, 3] | 5000 | 0.8925143953934741 |
| *SVC* | stemming | [3, 3] | 25000 | 0.9284343295859611 |