Train data set : 37,500개

Validation data set : 12,500개

1. 모든 Feature를 다 이용했을 때 (Feature의 수 : 7,996,039개)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.8696 | 0 | 0.88104 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.8696 | 0 | 0.8872 | 1 | Memory Error | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.88544 | 1 | 0.88488 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.88544 | 1 | 0.88928 | 1 | Memory Error | |

1. Max Feature를 1,000,000개로 했을 때

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.8796 | 1 | 0.87728 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.8796 | 1 | 0.88408 | 1 | Memory Error | |

1. Max Feature를 100,000개로 했을 때
2. Max Feature를 5,000개로 했을 때

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.85384 | 1 | 0.85216 | 1 | 0.81312 | 0 |
| Tfidf Vectorizer | 0.85384 | 1 | 0.85904 | 1 | 0.826 | 1 |