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 | 0.84112 | 0 | 0.87696 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.84112 | 0 | 0.87552 | 1 | Memory Error | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.88016 | 1 | 0.87992 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.88016 | 1 | 0.88056 | 1 | Memory Error | |

1. Max Feature를 1,000,000개로 했을 때
2. Max Feature를 500,000개로 했을 때

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.87992 | 1 | 0.87792 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.87992 | 1 | 0.8816 | 1 | Memory Error | |

1. Max Feature를 100,000개로 했을 때

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.87328 | 1 | 0.87152 | 1 | Memory Error | |
| Tfidf Vectorizer | 0.87328 | 1 | 0.87824 | 1 | Memory Error | |

1. Max Feature를 5,000개로 했을 때

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bernoulli NB | | Multinomial NB | | Gaussian NB | |
|  | Accuracy | Predict | Accuracy | Predict | Accuracy | Predict |
| Count Vectorizer | 0.84632 | 1 | 0.84168 | 0 | 0.80472 | 0 |
| Tfidf Vectorizer | 0.84632 | 1 | 0.84952 | 1 | 0.82024 | 1 |