the performance of original cleaned reviews in Sentiment analysis

original_clean_reviews=review_cleaner(train['review'],lemmatize=False,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews,
y=train["sentiment"],ngram=1,max_features=1000)

The training accuracy is: 0.99985 The validation accuracy is: 0.824

1. In **UNIGRAM setting** ie. when ngram=1 in the function train_predict_sentiment().Compare the performance of original cleaned reviews in Sentiment analysis to
    1. lemmatized reviews

original_clean_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews,
y=train["sentiment"],ngram=1,max_features=1000)

The training accuracy is: 1.0 The validation accuracy is: 0.8214

Compared to the performance of original cleaned reviews in Sentiment analysis, the training accuracy is higher but not validation accuracy

    2. stemmed reviews

original_clean_reviews=review_cleaner(train['review'],lemmatize=False,stem=True)

train_predict_sentiment(cleaned_reviews=original_clean_reviews,
y=train["sentiment"],ngram=1,max_features=1000)

The training accuracy is: 1.0 The validation accuracy is: 0.8266

Compared to the performance of original cleaned reviews in Sentiment analysis, the training accuracy and validation accuracy are higher

2. In **BIGRAM setting** ie. when ngram=2 in the function train_predict_sentiment().Compare the performance of original cleaned reviews in sentiment analysis to:
    1. lemmatized reviews

original_clean_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews,
y=train["sentiment"],ngram=2,max_features=1000)

The training accuracy is: 1.0 The validation accuracy is: 0.82

Compared to the performance of original cleaned reviews in Sentiment analysis, the training accuracy is higher but not validation accuracy

2. stemmed reviews

original_clean_reviews=review_cleaner(train['review'],lemmatize=False,stem=True)

train_predict_sentiment(cleaned_reviews=original_clean_reviews, y=train["sentiment"],ngram=2,max_features=1000)

The training accuracy is: 1.0 The validation accuracy is: 0.825

Compared to the performance of original cleaned reviews in Sentiment analysis, the training accuracy and validation accuracy are higher

3. In **UNIGRAM setting** and _lemmatize=True__ ie. when ngram=1, compare the performance of Sentiment analysis for these values of maximum features=[10,100,1000,5000], you can change the value of argument max_features in `train_predict_sentiment()

maximum feature = 10

original_clean_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews, y=train["sentiment"],ngram=1,max_features=10)

The training accuracy is: 0.8714 The validation accuracy is: 0.5638

maximum feature = 100

original_clean_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews, y=train["sentiment"],ngram=1,max_features=100)

The training accuracy is: 0.9999 The validation accuracy is: 0.7198

maximum feature = 1000

original_clean_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews, y=train["sentiment"],ngram=1,max_features=1000)

The training accuracy is: 0.99995 The validation accuracy is: 0.8204

maximum feature = 5000

original_clean_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)

train_predict_sentiment(cleaned_reviews=original_clean_reviews, y=train["sentiment"],ngram=1,max_features=5000)

The training accuracy is: 1.0 The validation accuracy is: 0.8452

Overall, when maximum feature is increasing and **UNIGRAM setting** is used , both the training accuracy and validation accuracy are increasing. Also, stemmed reviews give higher the training accuracy and validation accuracy regardless of argument value of ngram than lemmatized reviews and original cleaned reviews.