

# Dimitrios Hytioglou

## Unigram setting

### 1. Original

```
original_clean_lemmatized_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_reviews,
y=train["sentiment"],ngram=1,max_features=1000)
```

The training accuracy is: 0.99995  
The validation accuracy is: 0.8314

### 2. Lemmatized

```
original_clean_lemmatized_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_reviews,
y=train["sentiment"],ngram=1,max_features=1000)
```

The training accuracy is: 0.99995  
The validation accuracy is: 0.8314

### 3. Stemmed

```
original_clean_stemmed_reviews=review_cleaner(train['review'],lemmatize=False,stem=True)
train_predict_sentiment(cleaned_reviews=original_clean_stemmed_reviews,
y=train["sentiment"],ngram=1,max_features=1000)
```

The training accuracy is: 1.0  
The validation accuracy is: 0.819

## Bigram setting

### 1. Original

```
original_clean_reviews=review_cleaner(train['review'],lemmatize=False,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_reviews,
y=train["sentiment"],ngram=2,max_features=1000)
```

The training accuracy is: 0.99995  
The validation accuracy is: 0.8208

## 2. Lemmatized

```
original_clean_lemmatized_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_reviews,
y=train["sentiment"],ngram=2,max_features=1000)
```

The training accuracy is: 0.99995  
The validation accuracy is: 0.8184

## 3. Stemmed

```
original_clean_stemmed_reviews=review_cleaner(train['review'],lemmatize=False,stem=True)
train_predict_sentiment(cleaned_reviews=original_clean_stemmed_reviews,
y=train["sentiment"],ngram=2,max_features=1000)
```

The training accuracy is: 0.99995  
The validation accuracy is: 0.8216

# Comparison of various feature ranges in unigram setting

## 1. 10 features

```
original_clean_lemmatized_10_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_10_reviews,
y=train["sentiment"],ngram=1,max_features=10)
```

The training accuracy is: 0.87125  
The validation accuracy is: 0.5648

## 2. 100 features

```
original_clean_lemmatized_100_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_100_reviews,
y=train["sentiment"],ngram=1,max_features=100)
```

The training accuracy is: 0.99995  
The validation accuracy is: 0.7228

## 3. 1000 features

```
original_clean_lemmatized_1000_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_1000_reviews,
y=train["sentiment"],ngram=1,max_features=1000)
```

The training accuracy is: 1.0  
The validation accuracy is: 0.8256

## 4. 5000 features

```
original_clean_lemmatized_5000_reviews=review_cleaner(train['review'],lemmatize=True,stem=False)
train_predict_sentiment(cleaned_reviews=original_clean_lemmatized_5000_reviews,
y=train["sentiment"],ngram=1,max_features=5000)
```

The training accuracy is: 1.0

The validation accuracy is: 0.8394