NLP Analysis Report for imdb

text_length_stats

basic_stats

count   25000.00000

mean   1325.06964

std   1003.13367

min   52.00000

25%   702.00000

50%   979.00000

75%   1614.00000

max   13704.00000

Name: text_length, dtype: float64

skewness

2.162640484837894

kurtosis

6.5297628028330745

word_count_stats

basic_stats

count   25000.000000

mean   233.787200

std   173.733032

min   10.000000

25%   127.000000

```
50%      174.000000

75%      284.000000

max     2470.000000

Name: word_count, dtype: float64
```

vocabulary_diversity

```
0.7045722047760129
```

label_distribution

```
label

0    0.5

1    0.5

Name: proportion, dtype: float64
```

text_complexity

mean_complexity

```
0.20431511655110737
```

complexity_distribution

```
count    25000.000000

mean         0.204315

std          0.052611

min          0.013793

25%          0.168987

50%          0.202703

75%          0.237762

max          0.535714

Name: text_complexity, dtype: float64
```