SCRAPING DATA

[ ] ↳ 4 cells hidden

PREPROCESSING DATA

[ ] ↳ 35 cells hidden

LABELLING DATA LEXICON BASED

[ ] ↳ 7 cells hidden

MODEL ALGORITMA SUPPORT VECTOR MACHINE

[ ] ↳ 18 cells hidden

WORDCLOUD

↳ 6 cells hidden

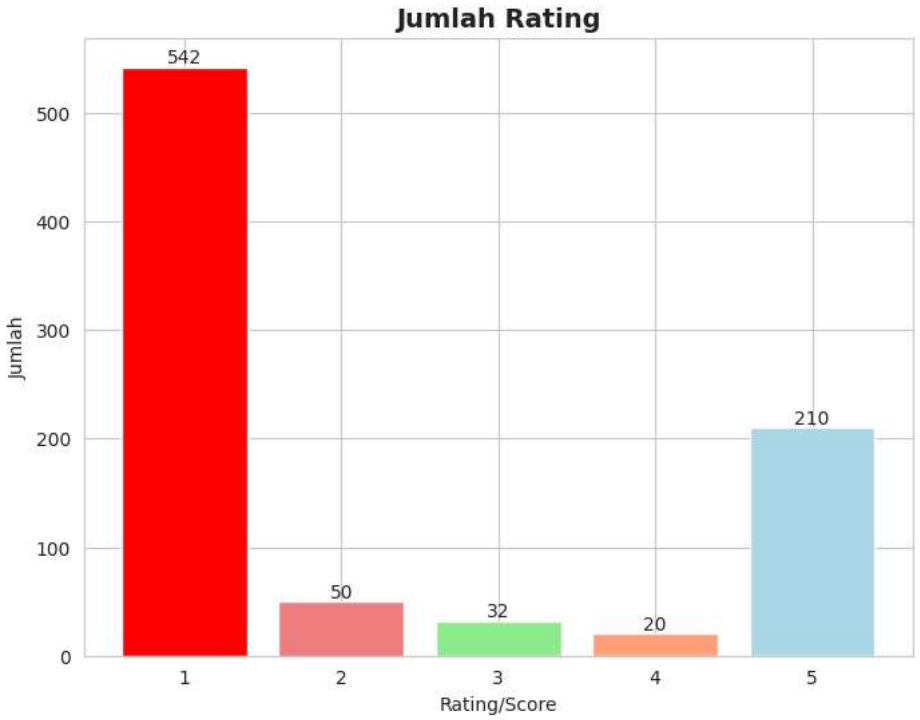
RATING

```
import pandas as pd

data = pd.read_csv('Hasil_Labelling_Data.csv')
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 854 entries, 0 to 853
Data columns (total 11 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Date              854 non-null    object
 1   Username          854 non-null    object
 2   Rating            854 non-null    int64 
 3   Review Text       854 non-null    object
 4   cleaning          854 non-null    object
 5   case folding      854 non-null    object
 6   normalisasi       854 non-null    object
 7   tokenize          854 non-null    object
 8   stopword removal  854 non-null    object
 9   steming_data      854 non-null    object
 10  Sentiment         854 non-null    object
dtypes: int64(1), object(10)
memory usage: 73.5+ KB
```

```
import matplotlib.pyplot as plt

rating_counts = data['Rating'].value_counts()
rating_counts = rating_counts.sort_index()

colors = ['red', 'lightcoral', 'lightgreen', 'lightsalmon', 'lightblue']

plt.figure(figsize=(8,6))
bars = plt.bar(rating_counts.index, rating_counts.values, color=colors)
plt.title('Jumlah Rating', fontsize=14, fontweight='bold')
plt.xlabel('Rating/Score')
plt.ylabel('Jumlah')
plt.xticks(rating_counts.index)

for bar in bars:
  height = bar.get_height()
  plt.text(bar.get_x() + bar.get_width() / 2, height, str(int(height)), ha='center', va='bottom')
```

```
plt. show()
```


Jumlah Rating
542
50
32
20
210
0
100
200
300
400
500
Jumlah
1
2
3
4
5
Rating/Score