RailwaySentimentAnalysis

June 24, 2025

1 Railway Sentiment Analysis

```python
[5]: import nltk
     nltk.download('punkt_tab') #tokenizer
```

```
[nltk_data] Downloading package punkt_tab to
[nltk_data]     /Users/kishankunal/nltk_data…
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
```

```
[5]: True
```

```python
[14]: import pandas as pd
      df  = pd.read_csv("./data/train.csv", encoding = "latin-1", header = None,␣
       ↪names = ['index', 'sentiment_label', 'text']) #kept data in same directory␣
       ↪under folder name data
      df.head()
```

```
[14]:       index sentiment_label                                          text
      0  Item ID         Sentiment                               SentimentText
      1        1                 0  @RailMinIndia My PNR is 8348062961, I am in wa…
      2        2                 0  @sureshpprabhu @RailMinIndia AC not working in…
      3        3                 0  @RailMinIndia I'm traveling to chennai by trai…
      4        4                 5  @RailMinIndia irctc is not responding at the t…
```

```python
[15]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1367 entries, 0 to 1366
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   index            1367 non-null   object
 1   sentiment_label  1367 non-null   object
 2   text             1367 non-null   object
dtypes: object(3)
memory usage: 32.2+ KB
```

```python
[16]: df.drop(0, inplace=True) #removing header as it is redundant
```

```
[19]:  #Displaying schema and null check
       print(df.info())
       print(df.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1366 entries, 1 to 1366
Data columns (total 3 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   index           1366 non-null   object
 1   sentiment_label 1366 non-null   object
 2   text            1366 non-null   object
dtypes: object(3)
memory usage: 32.1+ KB
None
index             0
sentiment_label   0
text              0
dtype: int64
```

```
[26]:  import matplotlib.pyplot as plt
       import seaborn as sns

       # describe dataset
       print(df.describe())

       # Distribution of sentiment labels
       plt.figure(figsize=(8,4))
       sns.countplot(x='sentiment_label', data=df)
       plt.title('Sentiment Label Distribution')
       plt.xlabel('Sentiment Class')
       plt.ylabel('Count')
       plt.show()

       # Tweet Length Distribution
       df['tweet_length'] = df['text'].apply(len)
       plt.figure(figsize=(8,4))
       sns.histplot(df['tweet_length'], bins = 30, kde=True)
       plt.title('Tweet Length Distribution')
       plt.xlabel('Character Length')
       plt.ylabel('Frequency')
       plt.show()
```
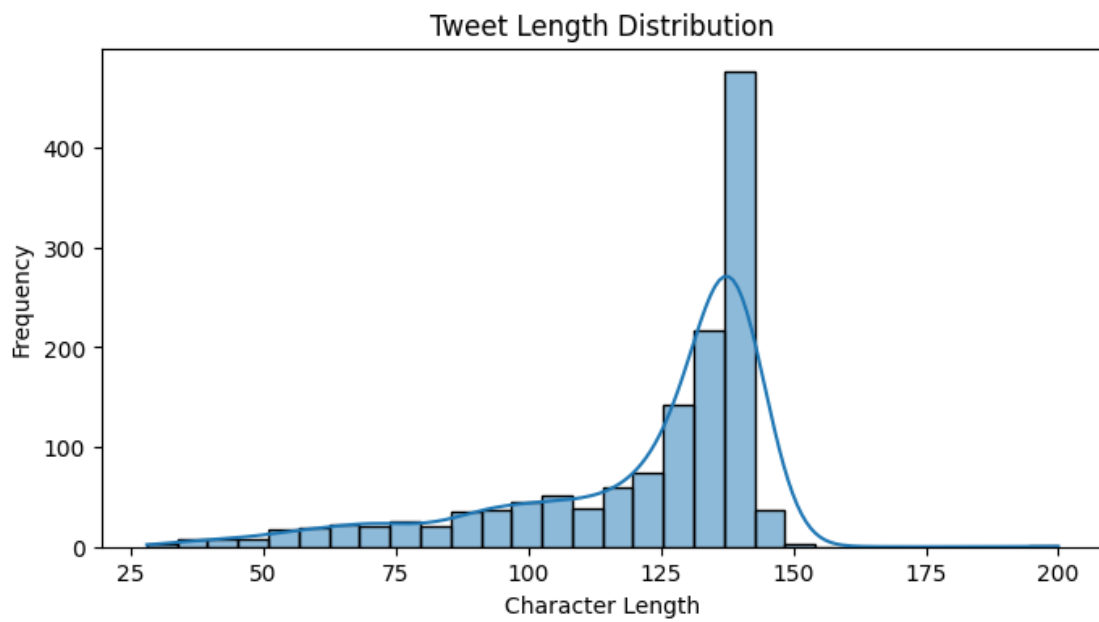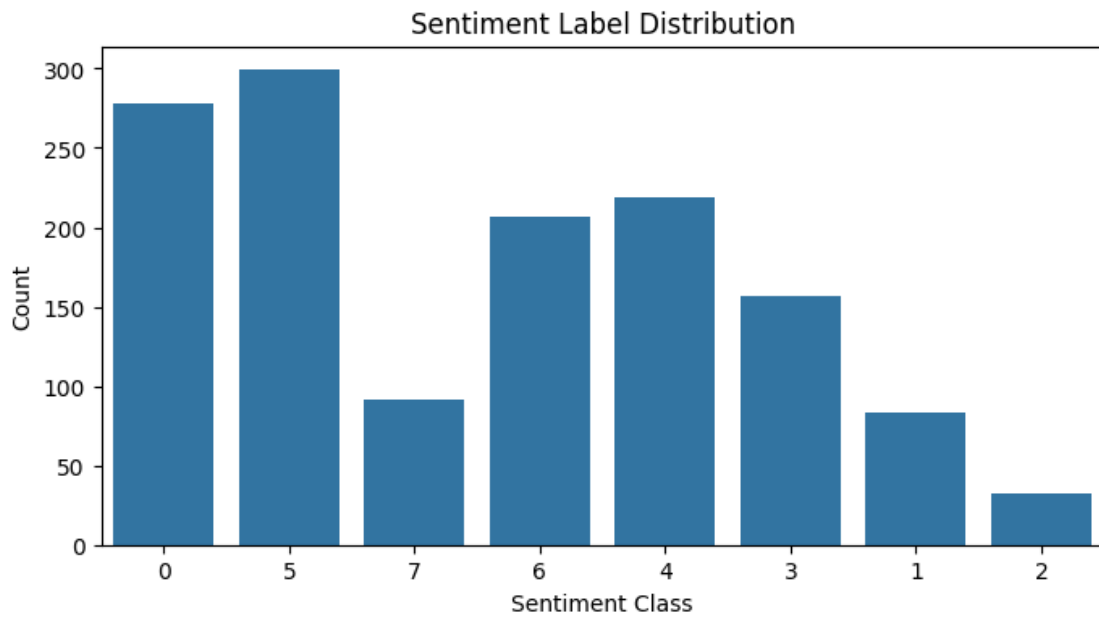
```
        index sentiment_label  \
count    1366            1366
unique   1366               8
top         1               5
freq        1             299
```

```
                                                    text
count                                               1366
unique                                              1353
top      @rpfwrbct @rpfwr1 @RailMinIndia @Deshbhakth1 @…
freq                                                   2
```


Sentiment Label Distribution

Tweet Length Distribution

- Sentiment labels range from 0 (Negative) to 7(Positive)
- A few outliers with unusually long lengths (~200+) likely due to urls or concateneated message

2 1.User-Mention Extraction

```
[28]: import re #regular expression library


      def extract_mentions(text):
          return re.findall(r'(@\w+)', text) #checking user handle @

      df['mentions'] = df['text'].apply(extract_mentions)
```

```
[29]: df.head()
```

```
[29]:    index  sentiment_label                                               text  \
       1      1                0  @RailMinIndia My PNR is 8348062961, I am in wa…
       2      2                0  @sureshpprabhu @RailMinIndia AC not working in…
       3      3                0  @RailMinIndia I'm traveling to chennai by trai…
       4      4                5  @RailMinIndia irctc is not responding at the t…
       5      5                7  @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…


          tweet_length                                           mentions
       1           140                                    [@RailMinIndia]
       2           137                     [@sureshpprabhu, @RailMinIndia]
       3           122                                    [@RailMinIndia]
       4            68                                    [@RailMinIndia]
       5           104  [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …
```

3 2. Hastag extraction

```
[35]: def extract_hastags(text):
          return [tag.lower() for tag in re.findall(r'#(\w+)', text)]

      df['hastags'] = df['text'].apply(extract_hastags)
      df['hastags'] = df['hastags'].apply(lambda x: list(set(x)))
      df.head(20)
```

```
[35]:    index  sentiment_label                                               text  \
       1      1                0  @RailMinIndia My PNR is 8348062961, I am in wa…
       2      2                0  @sureshpprabhu @RailMinIndia AC not working in…
       3      3                0  @RailMinIndia I'm traveling to chennai by trai…
       4      4                5  @RailMinIndia irctc is not responding at the t…
       5      5                7  @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
       6      6                6  @RailMinIndia If you can't give justice to gra…
       7      7                4  @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
       8      8                0  @RailMinIndia @mumbairailusers Dirty water flo…
```

```
9      9                    6    Enter to win $150 Amazon Gift Card! #Books #PN…
10     10                   3    @RailMinIndia still vendors are selling local …
11     11                   4    @Central_Railway @RailMinIndia @drmmumbaicr pl…
12     12                   0    @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
13     13                   5    @RailMinIndia @sureshpprabhu my reserved ticke…
14     14                   4    @RailMinIndia  update:- today intercity 12930 …
15     15                   0    @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
16     16                   6    The Google i1 button is used 5 billion times p…
17     17                   3    @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18     18                   4    @RailMinIndia Train has departed Varanasi and …
19     19                   4    @sureshpprabhu  @RailMinIndia student's pnr 21…
20     20                   5    @RailMinIndia Dear Sir, my father is a senior …
```

```
    tweet_length                                        mentions  \
1            140                                  [@RailMinIndia]
2            137                  [@sureshpprabhu, @RailMinIndia]
3            122                                  [@RailMinIndia]
4             68                                  [@RailMinIndia]
5            104    [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …
6            127                                  [@RailMinIndia]
7            139               [@sureshpprabhu, @RailMinIndiaPlz]
8            131                [@RailMinIndia, @mumbairailusers]
9             94                                               []
10           117                                  [@RailMinIndia]
11           120    [@Central_Railway, @RailMinIndia, @drmmumbaicr]
12           140                                  [@RailMinIndia]
13           137                  [@RailMinIndia, @sureshpprabhu]
14            58                                  [@RailMinIndia]
15           121                                  [@RailMinIndia]
16            74                                               []
17           140                                  [@RailMinIndia]
18           103                                  [@RailMinIndia]
19           140                  [@sureshpprabhu, @RailMinIndia]
20           140                                  [@RailMinIndia]
```

```
                                 hastags
1                                     []
2                                     []
3                                     []
4                                     []
5                                     []
6                                     []
7                                     []
8                                     []
9     [romance, books, urbanfantasy, pnr]
10                                    []
11                                    []
```

```
12                           [no, swatch]
13                                    []
14                                    []
15                                    []
16                                    []
17                                    []
18                                    []
19                                    []
20                                    []
```

4 3.Extract urls

```python
[40]: def extract_urls(text):
          urls = re.findall(r'https?://\S+', text)
          return urls

      df['urls'] = df['text'].apply(extract_urls)
      df['clean_text'] = df['text'].apply(lambda x: re.sub(r'https?://\S+', '', x))
      df.head(20)
```

```
[40]:     index sentiment_label                                              text  \
      1       1                0  @RailMinIndia My PNR is 8348062961, I am in wa…
      2       2                0  @sureshpprabhu @RailMinIndia AC not working in…
      3       3                0  @RailMinIndia I'm traveling to chennai by trai…
      4       4                5  @RailMinIndia irctc is not responding at the t…
      5       5                7  @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
      6       6                6  @RailMinIndia If you can't give justice to gra…
      7       7                4  @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
      8       8                0  @RailMinIndia @mumbairailusers Dirty water flo…
      9       9                6  Enter to win $150 Amazon Gift Card! #Books #PN…
      10     10                3  @RailMinIndia still vendors are selling local …
      11     11                4  @Central_Railway @RailMinIndia @drmmumbaicr pl…
      12     12                0  @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
      13     13                5  @RailMinIndia @sureshpprabhu my reserved ticke…
      14     14                4  @RailMinIndia  update:- today intercity 12930 …
      15     15                0  @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
      16     16                6  The Google i1 button is used 5 billion times p…
      17     17                3  @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
      18     18                4  @RailMinIndia Train has departed Varanasi and …
      19     19                4  @sureshpprabhu  @RailMinIndia student's pnr 21…
      20     20                5  @RailMinIndia Dear Sir, my father is a senior …

          tweet_length                                      mentions  \
      1            140                             [@RailMinIndia]
      2            137              [@sureshpprabhu, @RailMinIndia]
      3            122                             [@RailMinIndia]
```

```
4              68                          [@RailMinIndia]
5             104  [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …
6             127                          [@RailMinIndia]
7             139           [@sureshpprabhu, @RailMinIndiaPlz]
8             131            [@RailMinIndia, @mumbairailusers]
9              94                                       []
10            117                          [@RailMinIndia]
11            120   [@Central_Railway, @RailMinIndia, @drmmumbaicr]
12            140                          [@RailMinIndia]
13            137           [@RailMinIndia, @sureshpprabhu]
14             58                          [@RailMinIndia]
15            121                          [@RailMinIndia]
16             74                                       []
17            140                          [@RailMinIndia]
18            103                          [@RailMinIndia]
19            140           [@sureshpprabhu, @RailMinIndia]
20            140                          [@RailMinIndia]

                           hastags                         urls  \
1                              []                            []
2                              []                            []
3                              []                            []
4                              []                            []
5                              []                            []
6                              []                            []
7                              []                            []
8                              []    [https://t.co/mBZUOFkLXQ]
9    [romance, books, urbanfantasy, pnr]    [https://t.co/jBfXzDe7P2]
10                             []    [https://t.co/wdFyLlmMek]
11                             []                            []
12                  [no, swatch]    [https://t.co/h5uUdJuvel]
13                             []                            []
14                             []                            []
15                             []    [https://t.co/VIQmURi8HK]
16                             []                            []
17                             []    [https://t.co/sMfHUZdN2U]
18                             []                            []
19                             []                            []
20                             []                            []

                           clean_text
1   @RailMinIndia My PNR is 8348062961, I am in wa…
2   @sureshpprabhu @RailMinIndia AC not working in…
3   @RailMinIndia I'm traveling to chennai by trai…
4   @RailMinIndia irctc is not responding at the t…
5   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
6   @RailMinIndia If you can't give justice to gra…
```

```
7    @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8    @RailMinIndia @mumbairailusers Dirty water flo…
9    Enter to win $150 Amazon Gift Card! #Books #PN…
10   @RailMinIndia still vendors are selling local …
11   @Central_Railway @RailMinIndia @drmmumbaicr pl…
12   @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
13   @RailMinIndia @sureshpprabhu my reserved ticke…
14   @RailMinIndia  update:- today intercity 12930 …
15   @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
16   The Google i1 button is used 5 billion times p…
17   @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18   @RailMinIndia Train has departed Varanasi and …
19   @sureshpprabhu  @RailMinIndia student's pnr 21…
20   @RailMinIndia Dear Sir, my father is a senior …
```

5   4. extracting emojis

```python
import emoji

def replace_emoji(text):
    return emoji.demojize(text, delimiters=(" :", ": "))

df['clean_text'] = df['clean_text'].apply(replace_emoji)
df['clean_text'] = df['clean_text'].str.replace(r':[a-zA-Z0-9_]+:', '',
    regex=True)

df.head()
```

```
[43]:    index  sentiment_label                                               text  \
      1      1                0  @RailMinIndia My PNR is 8348062961, I am in wa…
      2      2                0  @sureshpprabhu @RailMinIndia AC not working in…
      3      3                0  @RailMinIndia I'm traveling to chennai by trai…
      4      4                5  @RailMinIndia irctc is not responding at the t…
      5      5                7  @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…

         tweet_length                                           mentions  hastags  \
      1           140                                    [@RailMinIndia]       []
      2           137                     [@sureshpprabhu, @RailMinIndia]       []
      3           122                                    [@RailMinIndia]       []
      4            68                                    [@RailMinIndia]       []
      5           104  [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …       []

        urls                                       clean_text
      1   []  @RailMinIndia My PNR is 8348062961, I am in wa…
      2   []  @sureshpprabhu @RailMinIndia AC not working in…
      3   []  @RailMinIndia I'm traveling to chennai by trai…
```

```
4   []   @RailMinIndia irctc is not responding at the t…
5   []   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
```

6 5. Text Pre-processing with NLTK & spaCy

```python
[48]: from nltk.tokenize import word_tokenize
      import spacy

      nlp = spacy.load('en_core_web_sm')

      #sample 100 random tweets

      sample_tweets = df.sample(n=100)['clean_text']

      nltk_tokens = sample_tweets.apply(word_tokenize)

      spacy_tokens = sample_tweets.apply(lambda x: [token.text for token in nlp(x)])

      #example comparison
      for i in range(5):
          print(f"\nTweet {i+1} :")
          print("NLTK : ", nltk_tokens.iloc[i])
          print("SpaCy : ", spacy_tokens.iloc[i])
```

```
Tweet 1 :
NLTK :  ['@', 'RailMinIndia', 'Railway', 'was', 'never', 'been', 'so',
'irresponsible', '.', 'so', 'called', 'fake', 'development', 'is', 'happening',
'.', 'railway', 'was', 'never', 'been', 'so', 'irresponsible', '.', '?', '?',
'?', '?']
SpaCy :  ['@RailMinIndia', 'Railway', 'was', 'never', 'been', 'so',
'irresponsible', '.', 'so', 'called', 'fake', 'development', 'is', 'happening',
'.', 'railway', 'was', 'never', 'been', 'so', 'irresponsible', '.', '?', '?',
'?', '?']

Tweet 2 :
NLTK :  ['@', 'leomanojc', '@', 'SirDebasis', '@', 'Naren_Reddy23', '@',
'SayNoToHindi', '@', 'mnsadhikrut', '@', 'abpmajhatv', '@', 'anilshidore', '@',
'RailMinIndiaâ\x80¦']
SpaCy :  ['@leomanojc', '@SirDebasis', '@Naren_Reddy23', '@SayNoToHindi',
'@mnsadhikrut', '@abpmajhatv', '@anilshidore', '@RailMinIndiaâ\x80', '¦']

Tweet 3 :
NLTK :  ['RT', '@', 'RailMinIndia', ':', 'Member', 'Traffic', 'reviewed', 'the',
'April-', 'Nov', 'transportation', 'performance', 'with', 'Rly', 'Board', 'and',
'Zonal', 'Rly', 'Officers', 'and', 'passedâ\x80¦']
SpaCy :  ['RT', '@RailMinIndia', ':', 'Member', 'Traffic', 'reviewed', 'the',
```

```
'April-', 'Nov', 'transportation', ' ', 'performance', 'with', 'Rly', 'Board',
'and', 'Zonal', 'Rly', 'Officers', 'and', 'passedâ\x80', '¦']

Tweet 4 :
NLTK : ['@', 'RailMinIndia', '@', 'sureshpprabhu', 'TRAIN', 'NO', '12764',
'S13', 'No', 'water', 'please', 'address', 'finding', 'difficult', 'Seat', 'no',
'S13', '42,43,45,46,48']
SpaCy : ['@RailMinIndia', '@sureshpprabhu', 'TRAIN', 'NO', '12764', 'S13',
'No', 'water', 'please', 'address', 'finding', 'difficult', 'Seat', 'no', 'S13',
'42,43,45,46,48']

Tweet 5 :
NLTK : ['@', 'RailMinIndia', 'my', 'pnr', 'no.2217356121', '.', 'pantry',
'charging', '20rs', 'for', 'water', 'bottle', 'which', 'has', 'printed', 'rate',
'of', '15']
SpaCy : ['@RailMinIndia', 'my', 'pnr', 'no.2217356121', '.', 'pantry',
'charging', '20rs', 'for', 'water', 'bottle', 'which', 'has', 'printed', 'rate',
'of', '15']
```

[52]:
```python
from nltk.stem import PorterStemmer

stemmer = PorterStemmer()

df['stemmed_text'] = df['clean_text'].apply(lambda x : ' '.join([stemmer.
 ↪stem(w) for w in x.split()]))
```

[58]:
```python
def lemmatize(text):
    doc = nlp(text)
    return ' '.join([token.lemma_ for token in doc])

df['lemmatized_text'] = df['clean_text'].apply(lemmatize)
```

[59]:
```python
def get_pos_tags(text):
    doc = nlp(text)
    return [(token.text, token.pos_, token.tag_) for token in doc]

df['pos_tags'] = df['clean_text'].apply(get_pos_tags)

df.head(20)
```

[59]:
```
   index sentiment_label                                              text  \
1      1               0  @RailMinIndia My PNR is 8348062961, I am in wa…
2      2               0  @sureshpprabhu @RailMinIndia AC not working in…
3      3               0  @RailMinIndia I'm traveling to chennai by trai…
4      4               5  @RailMinIndia irctc is not responding at the t…
5      5               7  @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
6      6               6  @RailMinIndia If you can't give justice to gra…
7      7               4  @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
```

```
8     8              0  @RailMinIndia @mumbairailusers Dirty water flo…
9     9              6  Enter to win $150 Amazon Gift Card! #Books #PN…
10    10             3  @RailMinIndia still vendors are selling local …
11    11             4  @Central_Railway @RailMinIndia @drmmumbaicr pl…
12    12             0  @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
13    13             5  @RailMinIndia @sureshpprabhu my reserved ticke…
14    14             4  @RailMinIndia  update:- today intercity 12930 …
15    15             0  @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
16    16             6  The Google i1 button is used 5 billion times p…
17    17             3  @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18    18             4  @RailMinIndia Train has departed Varanasi and …
19    19             4  @sureshpprabhu  @RailMinIndia student's pnr 21…
20    20             5  @RailMinIndia Dear Sir, my father is a senior …
```

```
     tweet_length                                           mentions  \
1             140                                    [@RailMinIndia]
2             137                      [@sureshpprabhu, @RailMinIndia]
3             122                                    [@RailMinIndia]
4              68                                    [@RailMinIndia]
5             104  [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …
6             127                                    [@RailMinIndia]
7             139                 [@sureshpprabhu, @RailMinIndiaPlz]
8             131                   [@RailMinIndia, @mumbairailusers]
9              94                                                 []
10            117                                    [@RailMinIndia]
11            120     [@Central_Railway, @RailMinIndia, @drmmumbaicr]
12            140                                    [@RailMinIndia]
13            137                [@RailMinIndia, @sureshpprabhu]
14             58                                    [@RailMinIndia]
15            121                                    [@RailMinIndia]
16             74                                                 []
17            140                                    [@RailMinIndia]
18            103                                    [@RailMinIndia]
19            140                 [@sureshpprabhu, @RailMinIndia]
20            140                                    [@RailMinIndia]
```

```
                            hastags                          urls  \
1                                []                            []
2                                []                            []
3                                []                            []
4                                []                            []
5                                []                            []
6                                []                            []
7                                []                            []
8                                []    [https://t.co/mBZUOFkLXQ]
9    [romance, books, urbanfantasy, pnr]    [https://t.co/jBfXzDe7P2]
10                               []    [https://t.co/wdFyLlmMek]
```

```
11                                          []                              []
12                             [no, swatch]  [https://t.co/h5uUdJuvel]
13                                          []                              []
14                                          []                              []
15                                          []  [https://t.co/VIQmURi8HK]
16                                          []                              []
17                                          []  [https://t.co/sMfHUZdN2U]
18                                          []                              []
19                                          []                              []
20                                          []                              []


                                         clean_text  \
1    @RailMinIndia My PNR is 8348062961, I am in wa…
2    @sureshpprabhu @RailMinIndia AC not working in…
3    @RailMinIndia I'm traveling to chennai by trai…
4    @RailMinIndia irctc is not responding at the t…
5    @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
6    @RailMinIndia If you can't give justice to gra…
7    @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8    @RailMinIndia @mumbairailusers Dirty water flo…
9    Enter to win $150 Amazon Gift Card! #Books #PN…
10   @RailMinIndia still vendors are selling local …
11   @Central_Railway @RailMinIndia @drmmumbaicr pl…
12   @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
13   @RailMinIndia @sureshpprabhu my reserved ticke…
14   @RailMinIndia  update:- today intercity 12930 …
15   @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
16   The Google i1 button is used 5 billion times p…
17   @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18   @RailMinIndia Train has departed Varanasi and …
19   @sureshpprabhu  @RailMinIndia student's pnr 21…
20   @RailMinIndia Dear Sir, my father is a senior …


                                       stemmed_text  \
1    @railminindia my pnr is 8348062961, i am in wa…
2    @sureshpprabhu @railminindia ac not work in pr…
3    @railminindia i'm travel to chennai by train 1…
4    @railminindia irctc is not respond at the time…
5    @drmbhopal @railminindia @sanjaygupta2012 @drm…
6    @railminindia if you can't give justic to grad…
7    @sureshpprabhu @railminindiaplz wrkout smthng …
8    @railminindia @mumbairailus dirti water flow f…
9    enter to win $150 amazon gift card! #book #pnr…
10   @railminindia still vendor are sell local loca…
11   @central_railway @railminindia @drmmumbaicr pl…
12   @railminindia #swatch bharat abhiyaan# train s…
13   @railminindia @sureshpprabhu my reserv ticket …
```

```
14  @railminindia update:- today interc 12930 late…
15  @railminindia train no 12656 there is no prope…
16  the googl i1 button is use 5 billion time pnr …
17  @railminindia sir, ?? ?????? ????? ?????? ?? ?…
18  @railminindia train ha depart varanasi and wil…
19  @sureshpprabhu @railminindia student' pnr 2119…
20  @railminindia dear sir, my father is a senior …


                                  lemmatized_text  \
1   @RailMinIndia my PNR be 8348062961 , I be in w…
2   @sureshpprabhu @railminindia AC not work in pr…
3   @railminindia I be travel to chennai by train …
4   @RailMinIndia irctc be not respond at the time…
5   @DRMbhopal @railminindia @sanjaygupta2012 @drm…
6   @RailMinIndia if you can not give justice to g…
7   @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8   @RailMinIndia @mumbairailusers Dirty water flo…
9   enter to win $ 150 Amazon Gift Card ! # Books …
10  @RailMinIndia still vendor be sell local local…
11  @central_railway @RailMinIndia @drmmumbaicr pl…
12  @RailMinIndia # Swatch Bharat Abhiyaan # Train…
13  @RailMinIndia @sureshpprabhu my reserved ticke…
14  @RailMinIndia   update:- today intercity 12930…
15  @RailMinIndia   TRAIN no 12656 there be no pro…
16  the Google i1 button be use 5 billion time pnr…
17  @RailMinIndia Sir , ? ? ? ? ? ? ? ? ? ? ? ? ? …
18  @RailMinIndia Train have depart Varanasi and w…
19  @sureshpprabhu   @RailMinIndia student 's pnr …
20  @RailMinIndia Dear Sir , my father be a senior…


                                          pos_tags
1   [(@RailMinIndia, NOUN, NNS), (My, PRON, PRP$),…
2   [(@sureshpprabhu, NOUN, NN), (@RailMinIndia, N…
3   [(@RailMinIndia, NOUN, NN), (I, PRON, PRP), ('…
4   [(@RailMinIndia, PROPN, NNP), (irctc, NOUN, NN…
5   [(@DRMbhopal, PROPN, NNP), (@RailMinIndia, NOU…
6   [(@RailMinIndia, NOUN, NNS), (If, SCONJ, IN), …
7   [(@sureshpprabhu, PROPN, NNP), (@RailMinIndiaP…
8   [(@RailMinIndia, PROPN, NNP), (@mumbairailuser…
9   [(Enter, VERB, VB), (to, PART, TO), (win, VERB…
10  [(@RailMinIndia, PROPN, NNP), (still, ADV, RB)…
11  [(@Central_Railway, NOUN, NN), (@RailMinIndia,…
12  [(@RailMinIndia, NOUN, NNS), (#, SYM, $), (Swa…
13  [(@RailMinIndia, NOUN, NNS), (@sureshpprabhu, …
14  [(@RailMinIndia, PROPN, NNP), ( , SPACE, _SP),…
15  [(@RailMinIndia, NOUN, NNS), ( , SPACE, _SP), …
16  [(The, DET, DT), (Google, PROPN, NNP), (i1, PR…
```

```
17   [(@RailMinIndia, PROPN, NNP), (Sir, PROPN, NNP…
18   [(@RailMinIndia, PROPN, NNP), (Train, PROPN, N…
19   [(@sureshpprabhu, NOUN, NN), ( , SPACE, _SP), …
20   [(@RailMinIndia, NOUN, NNS), (Dear, PROPN, NNP…
```

[60]: ```python
df.iloc[0]
```

[60]:
```
index   1
sentiment_label   0
text   @RailMinIndia My PNR is 8348062961, I am in wa…
tweet_length   140
mentions   [@RailMinIndia]
hastags   []
urls   []
clean_text   @RailMinIndia My PNR is 8348062961, I am in wa…
stemmed_text   @railminindia my pnr is 8348062961, i am in wa…
lemmatized_text   @RailMinIndia my PNR be 8348062961 , I be in w…
pos_tags   [(@RailMinIndia, NOUN, NNS), (My, PRON, PRP$),…
Name: 1, dtype: object
```

[62]: ```python
import time

start_time = time.time()

for _ in range(100):
    lemmatize(df.iloc[0]['clean_text'])
end_time = time.time()

avg_time_per_tweet = (end_time - start_time) / 100
print(f"Average processing time :  {avg_time_per_tweet : .4f} seconds")
```

```
Average processing time :   0.0067 seconds
```

7 6 NamedEntity Recognition

[64]: ```python
def extract_entites(text):
    doc = nlp(text)
    return [(ent.text, ent.label_) for ent in doc.ents]

df['entites'] = df['clean_text'].apply(extract_entites)
df.head()
```

[64]:
```
   index sentiment_label   text \
1   1   0 @RailMinIndia My PNR is 8348062961, I am in wa…
2   2   0 @sureshpprabhu @RailMinIndia AC not working in…
3   3   0 @RailMinIndia I'm traveling to chennai by trai…
4   4   5 @RailMinIndia irctc is not responding at the t…
```

```
5       5                      7  @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…

     tweet_length                                   mentions hastags  \
1             140                            [@RailMinIndia]      []
2             137              [@sureshpprabhu, @RailMinIndia]      []
3             122                            [@RailMinIndia]      []
4              68                            [@RailMinIndia]      []
5             104  [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …      []

  urls                                      clean_text  \
1   []   @RailMinIndia My PNR is 8348062961, I am in wa…
2   []   @sureshpprabhu @RailMinIndia AC not working in…
3   []   @RailMinIndia I'm traveling to chennai by trai…
4   []   @RailMinIndia irctc is not responding at the t…
5   []   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…

                                        stemmed_text  \
1  @railminindia my pnr is 8348062961, i am in wa…
2  @sureshpprabhu @railminindia ac not work in pr…
3  @railminindia i'm travel to chennai by train 1…
4  @railminindia irctc is not respond at the time…
5  @drmbhopal @railminindia @sanjaygupta2012 @drm…

                                     lemmatized_text  \
1  @RailMinIndia my PNR be 8348062961 , I be in w…
2  @sureshpprabhu @railminindia AC not work in pr…
3  @railminindia I be travel to chennai by train …
4  @RailMinIndia irctc be not respond at the time…
5  @DRMbhopal @railminindia @sanjaygupta2012 @drm…

                                            pos_tags  \
1  [(@RailMinIndia, NOUN, NNS), (My, PRON, PRP$),…
2  [(@sureshpprabhu, NOUN, NN), (@RailMinIndia, N…
3  [(@RailMinIndia, NOUN, NN), (I, PRON, PRP), ('…
4  [(@RailMinIndia, PROPN, NNP), (irctc, NOUN, NN…
5  [(@DRMbhopal, PROPN, NNP), (@RailMinIndia, NOU…

                                             entites
1                   [(PNR, ORG), (8348062961, DATE)]
2  [(@sureshpprabhu @RailMinIndia AC, ORG), (toda…
3  [(16102, CARDINAL), (S4-46, DATE), (one, CARDI…
4                                                 []
5                                                 []
```

```python
def map_domain_entites(entites):
    mapped = []
    for text, label in entites:
```

```
        if label == 'GPE':
            mapped.append((text, 'STATION'))
        elif label == 'CARDINAL' and len(text) > 2:
            mapped.append((text, 'TRAIN_NO'))
        elif label == 'ORDINAL':
            mapped.append((text, 'COACH'))
        elif text.lower in ['water', 'ac', 'cleanliness', 'toilet']:
            mapped.append((text, 'SERVICE'))
        else:
            mapped.append((text, 'OTHER'))

df['domain_entites'] = df['entites'].apply(map_domain_entites)

df.head(20)
```

[77]:   index sentiment_label   text \
    1   1   0   @RailMinIndia My PNR is 8348062961, I am in wa…
    2   2   0   @sureshpprabhu @RailMinIndia AC not working in…
    3   3   0   @RailMinIndia I'm traveling to chennai by trai…
    4   4   5   @RailMinIndia irctc is not responding at the t…
    5   5   7   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
    6   6   6   @RailMinIndia If you can't give justice to gra…
    7   7   4   @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
    8   8   0   @RailMinIndia @mumbairailusers Dirty water flo…
    9   9   6   Enter to win $150 Amazon Gift Card! #Books #PN…
    10   10   3   @RailMinIndia still vendors are selling local …
    11   11   4   @Central_Railway @RailMinIndia @drmmumbaicr pl…
    12   12   0   @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
    13   13   5   @RailMinIndia @sureshpprabhu my reserved ticke…
    14   14   4   @RailMinIndia update:- today intercity 12930 …
    15   15   0   @RailMinIndia TRAIN NO 12656 THERE IS NO PROP…
    16   16   6   The Google i1 button is used 5 billion times p…
    17   17   3   @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
    18   18   4   @RailMinIndia Train has departed Varanasi and …
    19   19   4   @sureshpprabhu @RailMinIndia student's pnr 21…
    20   20   5   @RailMinIndia Dear Sir, my father is a senior …

    tweet_length   mentions \
    1   140   [@RailMinIndia]
    2   137   [@sureshpprabhu, @RailMinIndia]
    3   122   [@RailMinIndia]
    4   68   [@RailMinIndia]
    5   104 [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …
    6   127   [@RailMinIndia]
    7   139   [@sureshpprabhu, @RailMinIndiaPlz]
    8   131   [@RailMinIndia, @mumbairailusers]
    9   94   []

16

```
10            117                              [@RailMinIndia]
11            120     [@Central_Railway, @RailMinIndia, @drmmumbaicr]
12            140                              [@RailMinIndia]
13            137                  [@RailMinIndia, @sureshpprabhu]
14             58                              [@RailMinIndia]
15            121                              [@RailMinIndia]
16             74                                           []
17            140                              [@RailMinIndia]
18            103                              [@RailMinIndia]
19            140                  [@sureshpprabhu, @RailMinIndia]
20            140                              [@RailMinIndia]


                              hastags                           urls  \
1                                  []                             []
2                                  []                             []
3                                  []                             []
4                                  []                             []
5                                  []                             []
6                                  []                             []
7                                  []                             []
8                                  []   [https://t.co/mBZUOFkLXQ]
9    [romance, books, urbanfantasy, pnr]  [https://t.co/jBfXzDe7P2]
10                                 []   [https://t.co/wdFyLlmMek]
11                                 []                             []
12                     [no, swatch]   [https://t.co/h5uUdJuvel]
13                                 []                             []
14                                 []                             []
15                                 []   [https://t.co/VIQmURi8HK]
16                                 []                             []
17                                 []   [https://t.co/sMfHUZdN2U]
18                                 []                             []
19                                 []                             []
20                                 []                             []


                                 clean_text  \
1    @RailMinIndia My PNR is 8348062961, I am in wa…
2    @sureshpprabhu @RailMinIndia AC not working in…
3    @RailMinIndia I'm traveling to chennai by trai…
4    @RailMinIndia irctc is not responding at the t…
5    @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
6    @RailMinIndia If you can't give justice to gra…
7    @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8    @RailMinIndia @mumbairailusers Dirty water flo…
9    Enter to win $150 Amazon Gift Card! #Books #PN…
10   @RailMinIndia still vendors are selling local …
11   @Central_Railway @RailMinIndia @drmmumbaicr pl…
12   @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
```

```
13  @RailMinIndia @sureshpprabhu my reserved ticke…
14  @RailMinIndia  update:- today intercity 12930 …
15  @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
16  The Google i1 button is used 5 billion times p…
17  @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18  @RailMinIndia Train has departed Varanasi and …
19  @sureshpprabhu  @RailMinIndia student's pnr 21…
20  @RailMinIndia Dear Sir, my father is a senior …


                                       stemmed_text  \
1   @railminindia my pnr is 8348062961, i am in wa…
2   @sureshpprabhu @railminindia ac not work in pr…
3   @railminindia i'm travel to chennai by train 1…
4   @railminindia irctc is not respond at the time…
5   @drmbhopal @railminindia @sanjaygupta2012 @drm…
6   @railminindia if you can't give justic to grad…
7   @sureshpprabhu @railminindiaplz wrkout smthng …
8   @railminindia @mumbairailus dirti water flow f…
9   enter to win $150 amazon gift card! #book #pnr…
10  @railminindia still vendor are sell local loca…
11  @central_railway @railminindia @drmmumbaicr pl…
12  @railminindia #swatch bharat abhiyaan# train s…
13  @railminindia @sureshpprabhu my reserv ticket …
14  @railminindia update:- today interc 12930 late…
15  @railminindia train no 12656 there is no prope…
16  the googl i1 button is use 5 billion time pnr …
17  @railminindia sir, ?? ?????? ????? ?????? ?? ?…
18  @railminindia train ha depart varanasi and wil…
19  @sureshpprabhu @railminindia student' pnr 2119…
20  @railminindia dear sir, my father is a senior …


                                    lemmatized_text  \
1   @RailMinIndia my PNR be 8348062961 , I be in w…
2   @sureshpprabhu @railminindia AC not work in pr…
3   @railminindia I be travel to chennai by train …
4   @RailMinIndia irctc be not respond at the time…
5   @DRMbhopal @railminindia @sanjaygupta2012 @drm…
6   @RailMinIndia if you can not give justice to g…
7   @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8   @RailMinIndia @mumbairailusers Dirty water flo…
9   enter to win $ 150 Amazon Gift Card ! # Books …
10  @RailMinIndia still vendor be sell local local…
11  @central_railway @RailMinIndia @drmmumbaicr pl…
12  @RailMinIndia # Swatch Bharat Abhiyaan # Train…
13  @RailMinIndia @sureshpprabhu my reserved ticke…
14  @RailMinIndia   update:- today intercity 12930…
15  @RailMinIndia   TRAIN no 12656 there be no pro…
```

```
16   the Google i1 button be use 5 billion time pnr…
17   @RailMinIndia Sir , ? ? ? ? ? ? ? ? ? ? ? ? ? …
18   @RailMinIndia Train have depart Varanasi and w…
19   @sureshpprabhu    @RailMinIndia student 's pnr …
20   @RailMinIndia Dear Sir , my father be a senior…
```

```
                                            pos_tags  \
1    [(@RailMinIndia, NOUN, NNS), (My, PRON, PRP$),…
2    [(@sureshpprabhu, NOUN, NN), (@RailMinIndia, N…
3    [(@RailMinIndia, NOUN, NN), (I, PRON, PRP), ('…
4    [(@RailMinIndia, PROPN, NNP), (irctc, NOUN, NN…
5    [(@DRMbhopal, PROPN, NNP), (@RailMinIndia, NOU…
6    [(@RailMinIndia, NOUN, NNS), (If, SCONJ, IN), …
7    [(@sureshpprabhu, PROPN, NNP), (@RailMinIndiaP…
8    [(@RailMinIndia, PROPN, NNP), (@mumbairailuser…
9    [(Enter, VERB, VB), (to, PART, TO), (win, VERB…
10   [(@RailMinIndia, PROPN, NNP), (still, ADV, RB)…
11   [(@Central_Railway, NOUN, NN), (@RailMinIndia,…
12   [(@RailMinIndia, NOUN, NNS), (#, SYM, $), (Swa…
13   [(@RailMinIndia, NOUN, NNS), (@sureshpprabhu, …
14   [(@RailMinIndia, PROPN, NNP), ( , SPACE, _SP),…
15   [(@RailMinIndia, NOUN, NNS), ( , SPACE, _SP), …
16   [(The, DET, DT), (Google, PROPN, NNP), (i1, PR…
17   [(@RailMinIndia, PROPN, NNP), (Sir, PROPN, NNP…
18   [(@RailMinIndia, PROPN, NNP), (Train, PROPN, N…
19   [(@sureshpprabhu, NOUN, NN), ( , SPACE, _SP), …
20   [(@RailMinIndia, NOUN, NNS), (Dear, PROPN, NNP…
```

```
                                        entites domain_entites
1                [(PNR, ORG), (8348062961, DATE)]          None
2    [(@sureshpprabhu @RailMinIndia AC, ORG), (toda…          None
3    [(16102, CARDINAL), (S4-46, DATE), (one, CARDI…          None
4                                              []          None
5                                              []          None
6                     [(one, CARDINAL), (SSE, ORG)]          None
7    [(@sureshpprabhu @RailMinIndiaPlz wrkout, ORG)…          None
8                            [(8.08, CARDINAL)]          None
9    [(150, MONEY), (Amazon, ORG), (Gift Card, PERS…          None
10                                             []          None
11   [(@Central_Railway @RailMinIndia, ORG), (12141…          None
12   [(Swatch Bharat Abhiyaan, ORG), (12218, DATE),…          None
13                                   [(RAC, ORG)]          None
14      [(today, DATE), (12930, DATE), (16, CARDINAL)]          None
15                                             []          None
16   [(Google, ORG), (5 billion, CARDINAL), (AllTwi…          None
17                                             []          None
18   [(@RailMinIndia Train, ORG), (Varanasi, LOC), …          None
```

```
19  [(@sureshpprabhu  @RailMinIndia, ORG), (211902…          None
20                                    [(RAC, ORG)]          None
```

```python
import pandas as pd
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer, negated
from afinn import Afinn
from nrclex import NRCLex

#initialing sentiment tools
vader_analyzer = SentimentIntensityAnalyzer()
afinn = Afinn()

def clean_text(text):
    return text.replace("RT", "").replace(r"@\S+", "").replace(r"#\S+","").
 ↪strip()

df['clean_text2'] = df['text'].apply(clean_text)

#compute sentiment scores

df['vader_score'] = df['clean_text2'].apply(lambda x : vader_analyzer.
 ↪polarity_scores(x)['compound'])
df['affin_score'] = df['clean_text2'].apply(lambda x : afinn.score(x))

def get_nrc_score(text):
    try:
        emotion = NRCLex(text = text)
        return sum(emotion.affect_frequencies.values())
    except Exception as e:
        return 0

df['nrc_score'] = df['clean_text'].apply(get_nrc_score)

threshold  = 0.05

df['vader_polarity'] = df['vader_score'].apply(lambda x : 'positive' if x >=␣
 ↪threshold else 'negative')
df['affin_polarity'] = df['affin_score'].apply(lambda x : 'positive' if x > 0␣
 ↪else 'negative')
df['nrc_polarity'] = df['nrc_score'].apply(lambda x : 'positive' if x > 0 else␣
 ↪'negative')


#compute alignment
df['alignment'] = df.apply(
    lambda row : 1 if row['vader_polarity'] == row['affin_polarity'] ==␣
 ↪row['nrc_polarity']  else 0,
```

```
    axis=1
)


#show some aligned vs non-aligned rows
aligned_df = df[df['alignment'] == 1]
non_aligned_df = df[df['alignment'] == 0]


# print result
total = len(df)
aligned = len(aligned_df)
agreement_rate = aligned / total*100

print(f"\nTotal Tweets : {total}")
print(f"Aligned Tweets : {aligned}")
print(f"Agreement rate : {agreement_rate : .2f}%")
```

```
Total Tweets : 1366
Aligned Tweets : 432
Agreement rate :  31.63%
```

[85]: `df.head()`

```
[85]:    index sentiment_label                                              text   \
       1     1               0   @RailMinIndia My PNR is 8348062961, I am in wa…
       2     2               0   @sureshpprabhu @RailMinIndia AC not working in…
       3     3               0   @RailMinIndia I'm traveling to chennai by trai…
       4     4               5   @RailMinIndia irctc is not responding at the t…
       5     5               7   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…

          tweet_length                                   mentions hastags   \
       1           140                             [@RailMinIndia]      []
       2           137            [@sureshpprabhu, @RailMinIndia]      []
       3           122                             [@RailMinIndia]      []
       4            68                             [@RailMinIndia]      []
       5           104   [@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …      []

         urls                                   clean_text   \
       1   []   @RailMinIndia My PNR is 8348062961, I am in wa…
       2   []   @sureshpprabhu @RailMinIndia AC not working in…
       3   []   @RailMinIndia I'm traveling to chennai by trai…
       4   []   @RailMinIndia irctc is not responding at the t…
       5   []   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…

                                      stemmed_text   \
       1   @railminindia my pnr is 8348062961, i am in wa…
```

```
2   @sureshpprabhu @railminindia ac not work in pr…
3   @railminindia i'm travel to chennai by train 1…
4   @railminindia irctc is not respond at the time…
5   @drmbhopal @railminindia @sanjaygupta2012 @drm…


                                   lemmatized_text  …  \
1   @RailMinIndia my PNR be 8348062961 , I be in w…  …
2   @sureshpprabhu @railminindia AC not work in pr…  …
3   @railminindia I be travel to chennai by train…   …
4   @RailMinIndia irctc be not respond at the time… …
5   @DRMbhopal @railminindia @sanjaygupta2012 @drm… …


                                            entites domain_entites  \
1                   [(PNR, ORG), (8348062961, DATE)]         None
2   [(@sureshpprabhu @RailMinIndia AC, ORG), (toda…         None
3   [(16102, CARDINAL), (S4-46, DATE), (one, CARDI…         None
4                                                []         None
5                                                []         None


                                        clean_text2 vader_score  affin_score  \
1   @RailMinIndia My PNR is 8348062961, I am in wa…     -0.4767         -2.0
2   @sureshpprabhu @RailMinIndia AC not working in…     -0.2235         -1.0
3   @RailMinIndia I'm traveling to chennai by trai…      0.0000         -2.0
4   @RailMinIndia irctc is not responding at the t…      0.0000          0.0
5   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…      0.0258         -1.0


    nrc_score  vader_polarity affin_polarity nrc_polarity alignment
1        1.0         negative       negative     positive         0
2        1.0         negative       negative     positive         0
3        1.0         negative       negative     positive         0
4        1.0         negative       negative     positive         0
5        1.0         negative       negative     positive         0

[5 rows x 21 columns]
```

```python
[96]:  #dependency chunk - liked/disliked clause extractor
       import spacy
       nlp = spacy.load('en_core_web_sm')

       def extract_clause(text, postive=True):
           doc = nlp(text)
           for sent in doc.sents:
               if postive and any(tok.lemma_ in {"good", "great", "thank"} for tok in
         ↪sent):
                   return sent.text.strip()
               if not postive and any(tok.lemma_ in {"bad", "not", "delay", "poor",
         ↪"dirty"} for tok in sent):
```

```
            return sent.text.strip()
```

```python
#filter 10 random positive and 10 negetive tweets
pos_tweets = df[df['vader_polarity'] == "positive"].sample(10, random_state=42)
neg_tweets = df[df['vader_polarity'] == "negative"].sample(10, random_state=42)

#extract liked/disliked clause

pos_tweets["liked_clause"] = pos_tweets['clean_text'].apply(lambda x :
  ↪extract_clause(x, True))
neg_tweets['disliked_clause'] = neg_tweets['clean_text'].apply(lambda x :
  ↪extract_clause(x, False))
```

[98]: `df.head(20)`

[98]:
	index	sentiment_label	text
1	1	0	@RailMinIndia My PNR is 8348062961, I am in wa…
2	2	0	@sureshpprabhu @RailMinIndia AC not working in…
3	3	0	@RailMinIndia I'm traveling to chennai by trai…
4	4	5	@RailMinIndia irctc is not responding at the t…
5	5	7	@DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
6	6	6	@RailMinIndia If you can't give justice to gra…
7	7	4	@sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8	8	0	@RailMinIndia @mumbairailusers Dirty water flo…
9	9	6	Enter to win $150 Amazon Gift Card! #Books #PN…
10	10	3	@RailMinIndia still vendors are selling local …
11	11	4	@Central_Railway @RailMinIndia @drmmumbaicr pl…
12	12	0	@RailMinIndia #Swatch Bharat Abhiyaan# Train s…
13	13	5	@RailMinIndia @sureshpprabhu my reserved ticke…
14	14	4	@RailMinIndia update:- today intercity 12930 …
15	15	0	@RailMinIndia TRAIN NO 12656 THERE IS NO PROP…
16	16	6	The Google i1 button is used 5 billion times p…
17	17	3	@RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18	18	4	@RailMinIndia Train has departed Varanasi and …
19	19	4	@sureshpprabhu @RailMinIndia student's pnr 21…
20	20	5	@RailMinIndia Dear Sir, my father is a senior …

	tweet_length	mentions
1	140	[@RailMinIndia]
2	137	[@sureshpprabhu, @RailMinIndia]
3	122	[@RailMinIndia]
4	68	[@RailMinIndia]
5	104	[@DRMbhopal, @RailMinIndia, @sanjaygupta2012, …
6	127	[@RailMinIndia]
7	139	[@sureshpprabhu, @RailMinIndiaPlz]
8	131	[@RailMinIndia, @mumbairailusers]
9	94	[]

```
10            117                                      [@RailMinIndia]
11            120      [@Central_Railway, @RailMinIndia, @drmmumbaicr]
12            140                                      [@RailMinIndia]
13            137                      [@RailMinIndia, @sureshpprabhu]
14             58                                      [@RailMinIndia]
15            121                                      [@RailMinIndia]
16             74                                                  []
17            140                                      [@RailMinIndia]
18            103                                      [@RailMinIndia]
19            140                      [@sureshpprabhu, @RailMinIndia]
20            140                                      [@RailMinIndia]

                            hastags                           urls  \
1                                []                             []
2                                []                             []
3                                []                             []
4                                []                             []
5                                []                             []
6                                []                             []
7                                []                             []
8                                []    [https://t.co/mBZUOFkLXQ]
9   [romance, books, urbanfantasy, pnr]    [https://t.co/jBfXzDe7P2]
10                               []    [https://t.co/wdFyLlmMek]
11                               []                             []
12                   [no, swatch]    [https://t.co/h5uUdJuvel]
13                               []                             []
14                               []                             []
15                               []    [https://t.co/VIQmURi8HK]
16                               []                             []
17                               []    [https://t.co/sMfHUZdN2U]
18                               []                             []
19                               []                             []
20                               []                             []

                            clean_text  \
1   @RailMinIndia My PNR is 8348062961, I am in wa…
2   @sureshpprabhu @RailMinIndia AC not working in…
3   @RailMinIndia I'm traveling to chennai by trai…
4   @RailMinIndia irctc is not responding at the t…
5   @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…
6   @RailMinIndia If you can't give justice to gra…
7   @sureshpprabhu @RailMinIndiaPlz wrkout smthng …
8   @RailMinIndia @mumbairailusers Dirty water flo…
9   Enter to win $150 Amazon Gift Card! #Books #PN…
10  @RailMinIndia still vendors are selling local …
11  @Central_Railway @RailMinIndia @drmmumbaicr pl…
12  @RailMinIndia #Swatch Bharat Abhiyaan# Train s…
```

```
13   @RailMinIndia @sureshpprabhu my reserved ticke…
14   @RailMinIndia  update:- today intercity 12930 …
15   @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…
16   The Google i1 button is used 5 billion times p…
17   @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…
18   @RailMinIndia Train has departed Varanasi and …
19   @sureshpprabhu  @RailMinIndia student's pnr 21…
20   @RailMinIndia Dear Sir, my father is a senior …


                                          stemmed_text  \
1    @railminindia my pnr is 8348062961, i am in wa…
2    @sureshpprabhu @railminindia ac not work in pr…
3    @railminindia i'm travel to chennai by train 1…
4    @railminindia irctc is not respond at the time…
5    @drmbhopal @railminindia @sanjaygupta2012 @drm…
6    @railminindia if you can't give justic to grad…
7    @sureshpprabhu @railminindiaplz wrkout smthng …
8    @railminindia @mumbairailus dirti water flow f…
9    enter to win $150 amazon gift card! #book #pnr…
10   @railminindia still vendor are sell local loca…
11   @central_railway @railminindia @drmmumbaicr pl…
12   @railminindia #swatch bharat abhiyaan# train s…
13   @railminindia @sureshpprabhu my reserv ticket …
14   @railminindia update:- today interc 12930 late…
15   @railminindia train no 12656 there is no prope…
16   the googl i1 button is use 5 billion time pnr …
17   @railminindia sir, ?? ?????? ????? ?????? ?? ?…
18   @railminindia train ha depart varanasi and wil…
19   @sureshpprabhu @railminindia student' pnr 2119…
20   @railminindia dear sir, my father is a senior …


                                       lemmatized_text  …  \
1    @RailMinIndia my PNR be 8348062961 , I be in w…  …
2    @sureshpprabhu @railminindia AC not work in pr…  …
3    @railminindia I be travel to chennai by train …  …
4    @RailMinIndia irctc be not respond at the time…  …
5    @DRMbhopal @railminindia @sanjaygupta2012 @drm…  …
6    @RailMinIndia if you can not give justice to g…  …
7    @sureshpprabhu @RailMinIndiaPlz wrkout smthng …  …
8    @RailMinIndia @mumbairailusers Dirty water flo…  …
9    enter to win $ 150 Amazon Gift Card ! # Books …  …
10   @RailMinIndia still vendor be sell local local…  …
11   @central_railway @RailMinIndia @drmmumbaicr pl…  …
12   @RailMinIndia # Swatch Bharat Abhiyaan # Train…  …
13   @RailMinIndia @sureshpprabhu my reserved ticke…  …
14   @RailMinIndia   update:- today intercity 12930…  …
15   @RailMinIndia   TRAIN no 12656 there be no pro…  …
```

```
16   the Google i1 button be use 5 billion time pnr…   …
17   @RailMinIndia Sir , ? ? ? ? ? ? ? ? ? ? ? ? ? …   …
18   @RailMinIndia Train have depart Varanasi and w…   …
19   @sureshpprabhu   @RailMinIndia student 's pnr …   …
20   @RailMinIndia Dear Sir , my father be a senior…   …
```

```
                                           clean_text2 vader_score affin_score  \
1    @RailMinIndia My PNR is 8348062961, I am in wa…      -0.4767        -2.0
2    @sureshpprabhu @RailMinIndia AC not working in…      -0.2235        -1.0
3    @RailMinIndia I'm traveling to chennai by trai…       0.0000        -2.0
4    @RailMinIndia irctc is not responding at the t…       0.0000         0.0
5    @DRMbhopal @RailMinIndia @sanjaygupta2012 @drm…       0.0258        -1.0
6    @RailMinIndia If you can't give justice to gra…      -0.6093         1.0
7    @sureshpprabhu @RailMinIndiaPlz wrkout smthng …      -0.2263        -1.0
8    @RailMinIndia @mumbairailusers Dirty water flo…      -0.4404        -2.0
9    Enter to win $150 Amazon Gift Card! #Books #PN…       0.9060         8.0
10   @RailMinIndia still vendors are selling local …       0.3612         2.0
11   @Central_Railway @RailMinIndia @drmmumbaicr pl…       0.3182         1.0
12   @RailMinIndia #Swatch Bharat Abhiyaan# Train s…      -0.4767        -3.0
13   @RailMinIndia @sureshpprabhu my reserved ticke…       0.8205         3.0
14   @RailMinIndia  update:- today intercity 12930 …       0.0000         0.0
15   @RailMinIndia  TRAIN NO 12656 THERE IS NO PROP…      -0.0258        -1.0
16   The Google i1 button is used 5 billion times p…       0.0000         0.0
17   @RailMinIndia Sir, ?? ?????? ????? ?????? ?? ?…       0.0000         0.0
18   @RailMinIndia Train has departed Varanasi and …       0.2023         1.0
19   @sureshpprabhu  @RailMinIndia student's pnr 21…      -0.3182        -1.0
20   @RailMinIndia Dear Sir, my father is a senior …       0.0772         0.0
```

```
   nrc_score  vader_polarity  affin_polarity  nrc_polarity alignment  \
1      1.0        negative        negative        positive        0
2      1.0        negative        negative        positive        0
3      1.0        negative        negative        positive        0
4      1.0        negative        negative        positive        0
5      1.0        negative        negative        positive        0
6      1.0        negative        positive        positive        0
7      1.0        negative        negative        positive        0
8      1.0        negative        negative        positive        0
9      0.0        positive        positive        negative        0
10     0.0        positive        positive        negative        0
11     1.0        positive        positive        positive        1
12     0.0        negative        negative        negative        1
13     1.0        positive        positive        positive        1
14     1.0        negative        negative        positive        0
15     0.0        negative        negative        negative        1
16     0.0        negative        negative        negative        1
17     0.0        negative        negative        negative        1
18     1.0        positive        positive        positive        1
```

```
19      1.0      negative      negative      positive      0
20      1.0      positive      negative      positive      0

    liked_clause disliked_clause
1            NaN             NaN
2            NaN             NaN
3            NaN             NaN
4            NaN             NaN
5            NaN             NaN
6            NaN             NaN
7            NaN             NaN
8            NaN             NaN
9            NaN             NaN
10           NaN             NaN
11           NaN             NaN
12           NaN             NaN
13           NaN             NaN
14           NaN             NaN
15           NaN             NaN
16           NaN             NaN
17           NaN             NaN
18           NaN             NaN
19           NaN             NaN
20           NaN             NaN

[20 rows x 23 columns]
```

8 image processing

```python
[103]: import os
import numpy as np
import matplotlib.pyplot as plt
from skimage import io, transform, filters, color, exposure
from pathlib import Path

def preprocess_images_skimage(data_dir, output_dir):
    """
    Preprocesses images in the specified data directory using scikit-image.
    Steps include: grayscale conversion, resizing, and histogram equalization.
    Processed images are saved to the output directory.

    Args:
        data_dir (str): Path to the input directory containing subfolders (e.g.
 ↪, '0', '1', ...)
                        which in turn contain emotion images (e.g., 'Happy.
 ↪jpg').
```

```python
        output_dir (str): Path to the directory where processed images will be
    ↪saved.
    """
    # Create output directory if it does not exist
    Path(output_dir).mkdir(parents=True, exist_ok=True)

    emotions = ['Anger', 'Contempt', 'Disgust', 'Fear', 'Happy', 'Sad',
    ↪'Surprised', 'Neutral']

    # Define target image size for resizing (e.g., 100x100 pixels)
    img_size = (100, 100)

    # Iterate through potential subject folders (numbered 0 to 18)
    for folder_num in range(19): # Assuming subject folders are named 0, 1, ...
    ↪, 18
        folder_path = os.path.join(data_dir, str(folder_num))

        # Check if the current path is actually a directory
        if not os.path.isdir(folder_path):
            # print(f"Skipping {folder_path} as it is not a directory.") #
    ↪Optional: uncomment for verbose skipping
            continue

        # Create corresponding output subdirectory for the current subject
    ↪folder
        output_folder_path = Path(output_dir) / str(folder_num)
        output_folder_path.mkdir(parents=True, exist_ok=True)

        # Process each emotion's image within the current subject folder
        for emotion in emotions:
            img_filename = f"{emotion}.jpg"
            img_path = os.path.join(folder_path, img_filename)

            # Check if the image file exists
            if not os.path.exists(img_path):
                # print(f"Image not found: {img_path}. Skipping.") # Optional:
    ↪uncomment for verbose skipping
                continue

            try:
                # 1. Load image
                img = io.imread(img_path)

                # 2. Convert to grayscale if it's a color image (handles RGB/
    ↪RGBA)
                if img.ndim == 3 and img.shape[2] >= 3:
```

```python
                img_gray = color.rgb2gray(img)
            else:
                # If it's already grayscale or single channel, use it as is.
                # Ensure it's float for subsequent operations if it wasn't
 ↪already.
                img_gray = img.astype(float)

            # 3. Resize image to a consistent size (e.g., 100x100)
            # anti_aliasing=True is recommended for downsampling to prevent
 ↪aliasing artifacts.
            img_resized = transform.resize(img_gray, img_size,
 ↪anti_aliasing=True)

            # 4. Apply Histogram Equalization for contrast enhancement
            # This operation typically outputs float values in the range [0.
 ↪0, 1.0].
            img_eq = exposure.equalize_hist(img_resized)

            # Convert the float image (0.0-1.0) to uint8 (0-255) for common
 ↪image formats like JPEG
            # Multiplying by 255 and casting to uint8 scales the pixel
 ↪values appropriately.
            img_to_save = (img_eq * 255).astype(np.uint8)

            # 5. Define the full path for the processed image in the output
 ↪directory
            output_img_path = output_folder_path / img_filename

            # 6. Save the processed image
            io.imsave(output_img_path, img_to_save)
            print(f"Processed and saved: {output_img_path}")

        except Exception as e:
            # Catch any errors during image processing (e.g., corrupted
 ↪files, unsupported formats)
            print(f"Error processing {img_path}: {e}")



input_data_directory = "data/images"  # Example: "C:/Users/YourUser/MyDataset"
 ↪or "/home/user/data/images"
output_results_directory = "processed_images"

print(f"Starting image preprocessing from: {os.path.
 ↪abspath(input_data_directory)}")
print(f"Output will be saved to: {os.path.abspath(output_results_directory)}")
```

29

```python
# Call the image preprocessing function
preprocess_images_skimage(input_data_directory, output_results_directory)

print("\nImage preprocessing task completed!")
```

Starting image preprocessing from: /Users/kishankunal/PycharmProjects/AppliedAI/
RailwayTwitterSentimentAnalysis/data/images
Output will be saved to: /Users/kishankunal/PycharmProjects/AppliedAI/RailwayTwi
tterSentimentAnalysis/processed_images
Processed and saved: processed_images/0/Anger.jpg
Processed and saved: processed_images/0/Contempt.jpg
Processed and saved: processed_images/0/Disgust.jpg
Processed and saved: processed_images/0/Fear.jpg
Processed and saved: processed_images/0/Happy.jpg
Processed and saved: processed_images/0/Sad.jpg
Processed and saved: processed_images/0/Surprised.jpg
Processed and saved: processed_images/0/Neutral.jpg
Processed and saved: processed_images/1/Anger.jpg
Processed and saved: processed_images/1/Contempt.jpg
Processed and saved: processed_images/1/Disgust.jpg
Processed and saved: processed_images/1/Fear.jpg
Processed and saved: processed_images/1/Happy.jpg
Processed and saved: processed_images/1/Sad.jpg
Processed and saved: processed_images/1/Surprised.jpg
Processed and saved: processed_images/1/Neutral.jpg
Processed and saved: processed_images/2/Anger.jpg
Processed and saved: processed_images/2/Contempt.jpg
Processed and saved: processed_images/2/Disgust.jpg
Processed and saved: processed_images/2/Fear.jpg
Processed and saved: processed_images/2/Happy.jpg
Processed and saved: processed_images/2/Sad.jpg
Processed and saved: processed_images/2/Surprised.jpg
Processed and saved: processed_images/2/Neutral.jpg
Processed and saved: processed_images/3/Anger.jpg
Processed and saved: processed_images/3/Contempt.jpg
Processed and saved: processed_images/3/Disgust.jpg
Processed and saved: processed_images/3/Fear.jpg
Processed and saved: processed_images/3/Happy.jpg
Processed and saved: processed_images/3/Sad.jpg
Processed and saved: processed_images/3/Surprised.jpg
Processed and saved: processed_images/3/Neutral.jpg
Processed and saved: processed_images/4/Anger.jpg
Processed and saved: processed_images/4/Contempt.jpg
Processed and saved: processed_images/4/Disgust.jpg
Processed and saved: processed_images/4/Fear.jpg
Processed and saved: processed_images/4/Happy.jpg
Processed and saved: processed_images/4/Sad.jpg

```
Processed and saved: processed_images/4/Surprised.jpg
Processed and saved: processed_images/4/Neutral.jpg
Processed and saved: processed_images/5/Anger.jpg
Processed and saved: processed_images/5/Contempt.jpg
Processed and saved: processed_images/5/Disgust.jpg
Processed and saved: processed_images/5/Fear.jpg
Processed and saved: processed_images/5/Happy.jpg
Processed and saved: processed_images/5/Sad.jpg
Processed and saved: processed_images/5/Surprised.jpg
Processed and saved: processed_images/5/Neutral.jpg
Processed and saved: processed_images/6/Anger.jpg
Processed and saved: processed_images/6/Contempt.jpg
Processed and saved: processed_images/6/Disgust.jpg
Processed and saved: processed_images/6/Fear.jpg
Processed and saved: processed_images/6/Happy.jpg
Processed and saved: processed_images/6/Sad.jpg
Processed and saved: processed_images/6/Surprised.jpg
Processed and saved: processed_images/6/Neutral.jpg
Processed and saved: processed_images/7/Anger.jpg
Processed and saved: processed_images/7/Contempt.jpg
Processed and saved: processed_images/7/Disgust.jpg
Processed and saved: processed_images/7/Fear.jpg
Processed and saved: processed_images/7/Happy.jpg
Processed and saved: processed_images/7/Sad.jpg
Processed and saved: processed_images/7/Surprised.jpg
Processed and saved: processed_images/7/Neutral.jpg
Processed and saved: processed_images/8/Anger.jpg
Processed and saved: processed_images/8/Contempt.jpg
Processed and saved: processed_images/8/Disgust.jpg
Processed and saved: processed_images/8/Fear.jpg
Processed and saved: processed_images/8/Happy.jpg
Processed and saved: processed_images/8/Sad.jpg
Processed and saved: processed_images/8/Surprised.jpg
Processed and saved: processed_images/8/Neutral.jpg
Processed and saved: processed_images/9/Anger.jpg
Processed and saved: processed_images/9/Contempt.jpg
Processed and saved: processed_images/9/Disgust.jpg
Processed and saved: processed_images/9/Fear.jpg
Processed and saved: processed_images/9/Happy.jpg
Processed and saved: processed_images/9/Sad.jpg
Processed and saved: processed_images/9/Surprised.jpg
Processed and saved: processed_images/9/Neutral.jpg
Processed and saved: processed_images/10/Anger.jpg
Processed and saved: processed_images/10/Contempt.jpg
Processed and saved: processed_images/10/Disgust.jpg
Processed and saved: processed_images/10/Fear.jpg
Processed and saved: processed_images/10/Happy.jpg
Processed and saved: processed_images/10/Sad.jpg
```

```
Processed and saved: processed_images/10/Surprised.jpg
Processed and saved: processed_images/10/Neutral.jpg
Processed and saved: processed_images/11/Anger.jpg
Processed and saved: processed_images/11/Contempt.jpg
Processed and saved: processed_images/11/Disgust.jpg
Processed and saved: processed_images/11/Fear.jpg
Processed and saved: processed_images/11/Happy.jpg
Processed and saved: processed_images/11/Sad.jpg
Processed and saved: processed_images/11/Surprised.jpg
Processed and saved: processed_images/11/Neutral.jpg
Processed and saved: processed_images/12/Anger.jpg
Processed and saved: processed_images/12/Contempt.jpg
Processed and saved: processed_images/12/Disgust.jpg
Processed and saved: processed_images/12/Fear.jpg
Processed and saved: processed_images/12/Happy.jpg
Processed and saved: processed_images/12/Sad.jpg
Processed and saved: processed_images/12/Surprised.jpg
Processed and saved: processed_images/12/Neutral.jpg
Processed and saved: processed_images/13/Anger.jpg
Processed and saved: processed_images/13/Contempt.jpg
Processed and saved: processed_images/13/Disgust.jpg
Processed and saved: processed_images/13/Fear.jpg
Processed and saved: processed_images/13/Happy.jpg
Processed and saved: processed_images/13/Sad.jpg
Processed and saved: processed_images/13/Surprised.jpg
Processed and saved: processed_images/13/Neutral.jpg
Processed and saved: processed_images/14/Anger.jpg
Processed and saved: processed_images/14/Contempt.jpg
Processed and saved: processed_images/14/Disgust.jpg
Processed and saved: processed_images/14/Fear.jpg
Processed and saved: processed_images/14/Happy.jpg
Processed and saved: processed_images/14/Sad.jpg
Processed and saved: processed_images/14/Surprised.jpg
Processed and saved: processed_images/14/Neutral.jpg
Processed and saved: processed_images/15/Anger.jpg
Processed and saved: processed_images/15/Contempt.jpg
Processed and saved: processed_images/15/Disgust.jpg
Processed and saved: processed_images/15/Fear.jpg
Processed and saved: processed_images/15/Happy.jpg
Processed and saved: processed_images/15/Sad.jpg
Processed and saved: processed_images/15/Surprised.jpg
Processed and saved: processed_images/15/Neutral.jpg
Processed and saved: processed_images/16/Anger.jpg
Processed and saved: processed_images/16/Contempt.jpg
Processed and saved: processed_images/16/Disgust.jpg
Processed and saved: processed_images/16/Fear.jpg
Processed and saved: processed_images/16/Happy.jpg
Processed and saved: processed_images/16/Sad.jpg
```

```
Processed and saved: processed_images/16/Surprised.jpg
Processed and saved: processed_images/16/Neutral.jpg
Processed and saved: processed_images/17/Anger.jpg
Processed and saved: processed_images/17/Contempt.jpg
Processed and saved: processed_images/17/Disgust.jpg
Processed and saved: processed_images/17/Fear.jpg
Processed and saved: processed_images/17/Happy.jpg
Processed and saved: processed_images/17/Sad.jpg
Processed and saved: processed_images/17/Surprised.jpg
Processed and saved: processed_images/17/Neutral.jpg
Processed and saved: processed_images/18/Anger.jpg
Processed and saved: processed_images/18/Contempt.jpg
Processed and saved: processed_images/18/Disgust.jpg
Processed and saved: processed_images/18/Fear.jpg
Processed and saved: processed_images/18/Happy.jpg
Processed and saved: processed_images/18/Sad.jpg
Processed and saved: processed_images/18/Surprised.jpg
Processed and saved: processed_images/18/Neutral.jpg

Image preprocessing task completed!
```

[ ]: