sentiment-analysis

June 3, 2023

#Sentiment analysis of Twitter layoffs

Importing libraries

[36]:
```
import pandas as pd
import numpy as np
import re
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

Importing dataset

[37]:
```
df = pd.read_excel('/content/layoffs.xlsx')
```

Describing dataset

[38]:
```
df.head()
```

[38]:

	Unnamed: 0.1	Unnamed: 0	Date	User	number of likes	source of tweet	count of replies	count of retweets	count of users that quoted the tweet and replied
0	0	0	2023-01-20	TXMCtrades	0	Twitter for iPhone	0	0	0
1	1	1	2023-01-20	otter25702	0	Twitter Web App	0	0	0
2	2	2	2023-01-20	Good_blessings	0	Twitter for Android	0	0	0
3	3	3	2023-01-20	Kbhatia98	0	Twitter Web App	0	0	0
4	4	4	2023-01-20	Calbert1994	0	Twitter Web App	0	0	0

```
2                                                    0
3                                                    0
4                                                    0

                                                Tweet language  \
0  Layoffs creating bullish PA $GOOGL https://t.c…       en
1  The market crash of 2008, where I lost my pens…       en
2  I'm just scared of layoffs. People can't imagi…       en
3  statutory warning: DON'T OPEN LINKEDIN FOR A M…       en
4        Tech industry layoffs feel like a grim omen       en

                                                media mentioned user in tweet  \
0  Photo(previewUrl='https://pbs.twimg.com/media/…                      NaN
1                                                NaN                      NaN
2                                                NaN                      NaN
3                                                NaN                      NaN
4                                                NaN                      NaN

  media_shared
0        Photo
1          NaN
2          NaN
3          NaN
4          NaN
```

[39]:
```
df.nunique()
```

[39]:
```
Unnamed: 0.1                                        30000
Unnamed: 0                                          30000
Date                                                    7
User                                                20871
number of likes                                       368
source of tweet                                       592
count of replies                                       82
count of retweets                                     135
count of users that quoted the tweet and replied       54
Tweet                                               29509
language                                               51
media                                                3894
mentioned user in tweet                              6543
media_shared                                            3
dtype: int64
```

[40]:
```
df.shape
```

[40]:
```
(30000, 14)
```

1. Extracting english tweets

[41]:
```
df = df[df['language'] == 'en']
df.shape
```

[41]:
```
(27482, 14)
```

2. Performing basic EDA

[42]:
```
print("First few rows:\n", df.head())
print("Summary statistics:\n", df.describe())

print("Data types:\n", df.dtypes)
print("Sentiment distribution:\n", df['Tweet'].value_counts())
```

```
First few rows:
    Unnamed: 0.1  Unnamed: 0       Date            User  number of likes  \
0             0           0 2023-01-20      TXMCtrades                0
1             1           1 2023-01-20      otter25702                0
2             2           2 2023-01-20  Good_blessings                0
3             3           3 2023-01-20       Kbhatia98                0
4             4           4 2023-01-20     Calbert1994                0

       source of tweet  count of replies  count of retweets  \
0   Twitter for iPhone                 0                  0
1      Twitter Web App                 0                  0
2  Twitter for Android                 0                  0
3      Twitter Web App                 0                  0
4      Twitter Web App                 0                  0

   count of users that quoted the tweet and replied  \
0                                                 0
1                                                 0
2                                                 0
3                                                 0
4                                                 0

                                              Tweet language  \
0  Layoffs creating bullish PA $GOOGL https://t.c…       en
1  The market crash of 2008, where I lost my pens…       en
2  I'm just scared of layoffs. People can't imagi…       en
3  statutory warning: DON'T OPEN LINKEDIN FOR A M…       en
4        Tech industry layoffs feel like a grim omen       en

                                              media mentioned user in tweet  \
0  Photo(previewUrl='https://pbs.twimg.com/media/…                      NaN
1                                                NaN                    NaN
2                                                NaN                    NaN
```

```
3                                                   NaN                    NaN
4                                                   NaN                    NaN

  media_shared
0        Photo
1          NaN
2          NaN
3          NaN
4          NaN
Summary statistics:
        Unnamed: 0.1    Unnamed: 0  number of likes  count of replies  \
count  27482.000000  27482.000000     27482.000000      27482.000000
mean   14881.012481  14881.012481        10.306382          1.108835
std     8624.543842   8624.543842       198.765611         43.093332
min        0.000000      0.000000         0.000000          0.000000
25%     7415.250000   7415.250000         0.000000          0.000000
50%    14757.500000  14757.500000         0.000000          0.000000
75%    22282.750000  22282.750000         2.000000          0.000000
max    29999.000000  29999.000000     20458.000000       5840.000000

       count of retweets  count of users that quoted the tweet and replied
count       27482.000000                                     27482.000000
mean            1.540099                                         0.237246
std            32.313971                                         6.440190
min             0.000000                                         0.000000
25%             0.000000                                         0.000000
50%             0.000000                                         0.000000
75%             0.000000                                         0.000000
max          3877.000000                                       812.000000
Data types:
 Unnamed: 0.1                                                    int64
Unnamed: 0                                                       int64
Date                                                    datetime64[ns]
User                                                            object
number of likes                                                  int64
source of tweet                                                 object
count of replies                                                 int64
count of retweets                                                int64
count of users that quoted the tweet and replied                 int64
Tweet                                                           object
language                                                        object
media                                                           object
mentioned user in tweet                                         object
media_shared                                                    object
dtype: object
Sentiment distribution:
 The bitcoin sector appears to have been hit by the wave of layoffs. A
cryptocurrency company, is planning to cut its personnel by nearly 20% after
```

```
coinbase. The corporation refused to say which department will be most affected.
It attribute                                                    10
Co-founder and ceo of kris marszalek, has announced a new wave of staff layoffs
that will reduce its global workforce by another 20%. Citing poor market
conditions and "recent industry events, this the right decision for the company
at this                                                         9
BUSINESS INSIDER: Facebook is conducting 'quiet layoffs' by urging managers to
label a certain number of workers as underperforming. The moves may lead to
thousands of job cuts. (DSP: AS PREDICTED BY THE TRILLION DOLLAR MAN!!)
7
The company laid off about 200 employees last week as part of a company-wide
restructuring. The mass layoffs would mark the second staff reduction soundhound
has undertaken in less than a year in november, the company reportedly laid off
10                                                              7
The supply chain crisis has had a profound impact across industries and
throughout the global economy. It has contributed to surging prices, layoffs,
productivity declines and empty store shelves.
6
                                                               ..
@Denzil101010 @IzzardBecky Yeah I understand… but the same financials also
affects the organization who have locked in contracts with player wages ! Loss
of revenue may mean layoffs to supporting staff in the long run. Or we just
expect the Shiek to inject more capital  as charity ?     1
Aaj ka gyaan : Layoffs ke next day Linkedin and Valentine ke next day Instagram
nhi kholna chaiye!!
1
#Headlines | Mid Day Prime\n\n Amazon begins its largest-ever layoffs\n Jacinda
Ardern to resign as New Zealand PM\n U.S. announces USD 125 M package for
Ukraine\n&amp;amp; more…\n\nWatch 'Mid Day Prime' for all the latest updates:
https://t.co/VH8VpPLiSn https://t.co/mknABxhlBx           1
@WOOLLEY_GAMER @DestinLegarie Destin makes a good point here. We have MS
attempting to buy up everything, making layoffs, and still delivering NO games.
If I was an Xbox only gamer I'd be extremely fed up.
1
Business is deteriorating rapidly at used-car seller Carvana, which is cutting
hours for its staff and big rounds of layoffs. The slowdown will make it harder
for the company to keep servicing its heavy debt load. https://t.co/dLL9UfPMVZ
1
Name: Tweet, Length: 27056, dtype: int64
```

3. Preprocessing

[46]:
```python
def preprocess_text(text):
    # Removing URLs
    text = re.sub(r'http\S+', '', text)
    # Removing mentions
    text = re.sub('@[^\s]+', '', text)
    # Removing special characters
```

```
    text = re.sub('[^a-zA-Z]', ' ', text)
    # Converting text to lowercase
    text = text.lower()
    # Removing stopwords
    text = ' '.join([word for word in text.split() if word not in
                     set(stopwords.words('english'))])
    return text

df['clean_tweet'] = df['Tweet'].apply(preprocess_text)
```

4. Transforming the dataset using TF-IDF vectorizer

[ ]:
```
vectorizer = TfidfVectorizer()

tfidf_matrix = vectorizer.fit_transform(df['clean_tweet'])
```

[47]:
```
tfidf_matrix
```

[47]:
```
<27482x25962 sparse matrix of type '<class 'numpy.float64'>'
        with 380285 stored elements in Compressed Sparse Row format>
```

[49]:
```
nltk.download('vader_lexicon')
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data…
```

[49]:
```
True
```

5. Evaluation using VADER model

[50]:
```
def evaluate_sentiment(text):
    analyzer = SentimentIntensityAnalyzer()
    sentiment_scores = analyzer.polarity_scores(text)
    return sentiment_scores['compound']

df['sentiment_score'] = df['Tweet'].apply(evaluate_sentiment)
```

[51]:
```
df['sentiment_score']
```

[51]:
```
0        0.2960
1       -0.3612
2       -0.0000
3       -0.3400
4       -0.2960
          …
29993    0.0000
29995    0.4019
29997    0.8797
29998   -0.0018
```

```
29999   -0.4588
Name: sentiment_score, Length: 27482, dtype: float64
```

6. Computing the percentage of positive and negative tweets

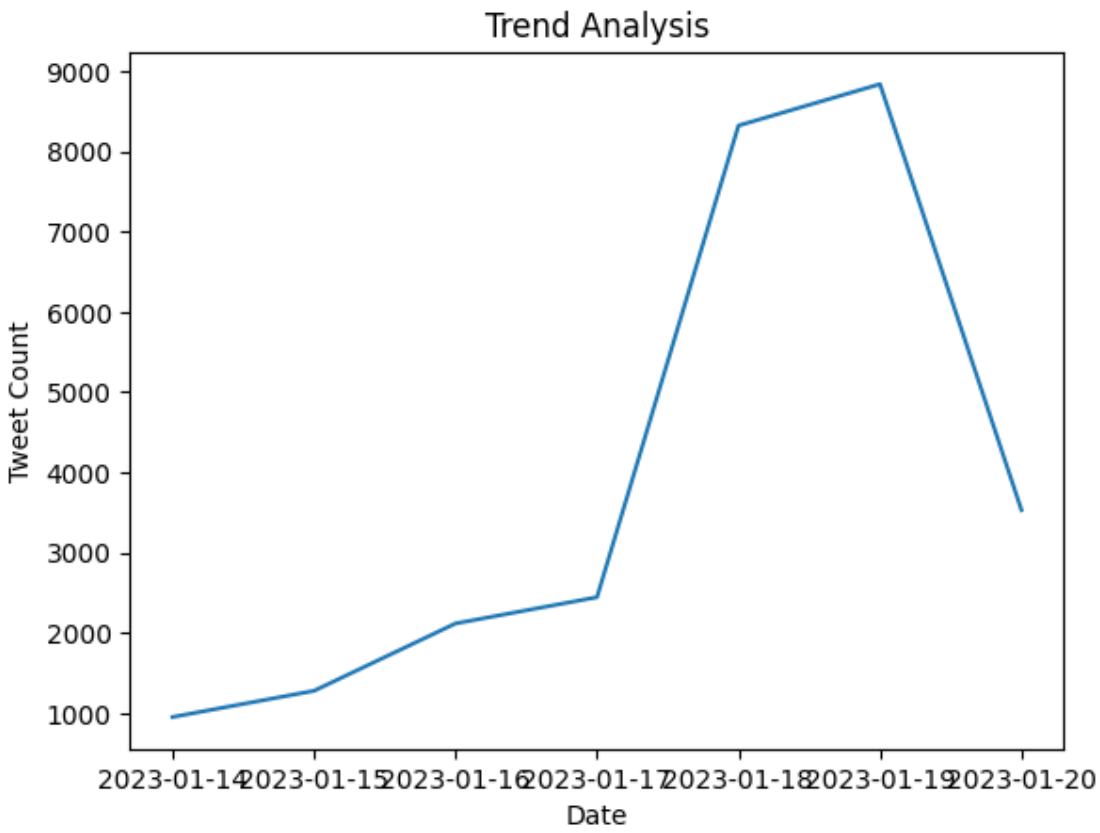
[83]:
```
positive_tweets = df[df['sentiment_score'] > 0.05]
negative_tweets = df[df['sentiment_score'] < -0.05]

positive_percentage = (len(positive_tweets) / len(df)) * 100
negative_percentage = (len(negative_tweets) / len(df)) * 100

print("Positive tweets count:", len(positive_tweets))
print("Negative tweets count:", len(negative_tweets))
print("Positive tweets percentage:", positive_percentage)
print("Negative tweets percentage:", negative_percentage)
```

```
Positive tweets count: 9239
Negative tweets count: 18243
Positive tweets percentage: 33.61836838658031
Negative tweets percentage: 66.38163161341969
```

7. Trend analysis by time series

[64]:
```
df['Date'] = pd.to_datetime(df['Date'])
df.set_index('Date', inplace = True)

daily_counts = df['Tweet'].resample('D').count()

plt.plot(daily_counts)
plt.xlabel('Date')
plt.ylabel('Tweet Count')
plt.title('Trend Analysis')
plt.show
```

[64]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



8. Calculating the frequency of hashtags and keywords

[68]:
```
hashtags = []
keywords = []

for tweet in df['Tweet']:
  # Hashtags
    hashtags.extend(re.findall(r'#(\w+)', tweet))
  # Keywords
    keywords.extend(re.findall(r'@(\w+)', tweet))

hashtags_freq = nltk.FreqDist(hashtags)
keywords_freq = nltk.FreqDist(keywords)
```

9. Top hashtags and keywords

[69]:
```
top_hashtags = hashtags_freq.most_common(10)
top_keywords = keywords_freq.most_common(10)

print("Top hashtags:", top_hashtags)
print("Top keywords:", top_keywords)
```

```
Top hashtags: [('layoffs', 4256), ('layoff', 1729), ('thelayoff', 1166),
('jobcuts', 1075), ('rifs', 763), ('Microsoft', 679), ('Layoffs', 574), ('jobs',
398), ('recession', 379), ('Amazon', 292)]
Top keywords: [('POTUS', 327), ('Microsoft', 217), ('WhiteHouse', 172),
('JayDubcity16', 127), ('elonmusk', 121), ('Jim_Jordan', 108), ('amazon', 103),
('RepJeffries', 97), ('jasonschreier', 85), ('pulte', 83)]
```

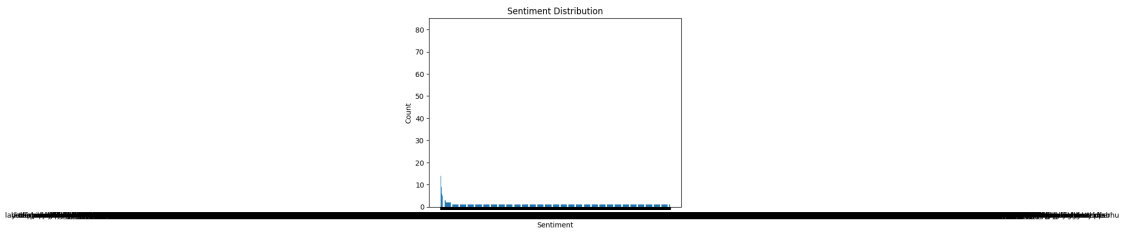
10. Identifying top users or influencers

[70]:
```
top_users = df['User'].value_counts().head(10)
print("Top users:\n", top_users)
```

```
Top users:
 TheLayoffBot       1479
jais_george           41
TickerTick_com        39
Gyan_Ganga_book       34
layoffs_              34
GetOnTrade            32
techinjektion         30
itsrohitchouhan       28
LayoffsTracker        28
moneycontrolcom       27
Name: User, dtype: int64
```

11. Data visualization of target variables

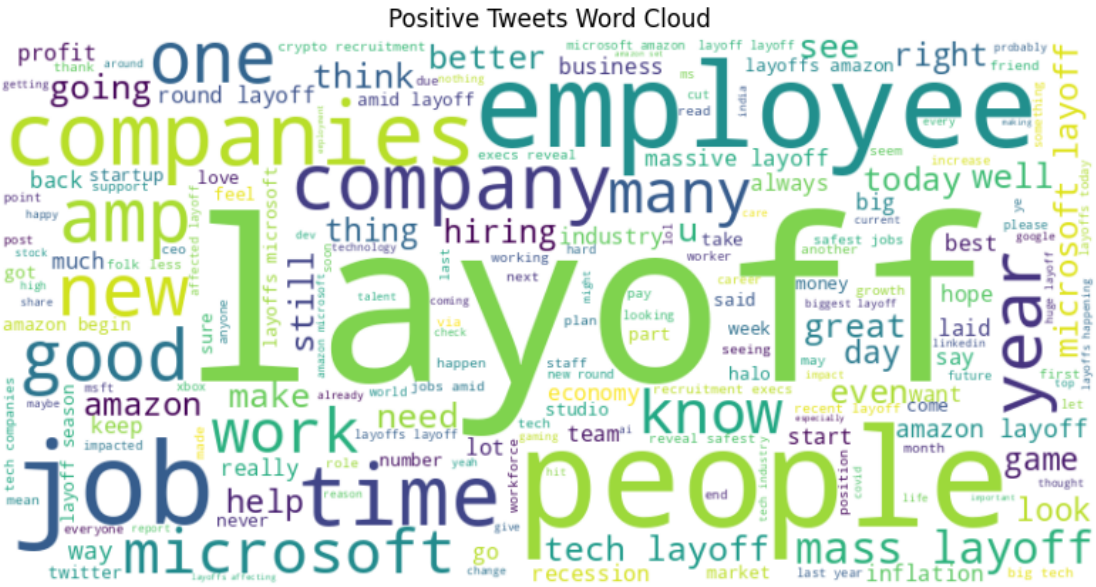
[72]:
```
sentiment_counts = df['clean_tweet'].value_counts()

plt.bar(sentiment_counts.index, sentiment_counts.values)
plt.xlabel('Sentiment')
plt.ylabel('Count')
plt.title('Sentiment Distribution')
plt.show()
```



12. Word cloud for positive and negative tweets

[85]:
```
positive_text = " ".join(positive_tweets['clean_tweet'])
positive_wordcloud = WordCloud(width = 800, height = 400,
                               background_color = 'white').generate(positive_text)
```

```
plt.figure(figsize=(10, 5))
plt.imshow(positive_wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Positive Tweets Word Cloud')
plt.show()
```



[86]:
```
negative_text = " ".join(negative_tweets['clean_tweet'])
negative_wordcloud = WordCloud(width = 800, height = 400,
                               background_color='white').generate(negative_text)

plt.figure(figsize=(10, 5))
plt.imshow(negative_wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Negative Tweets Word Cloud')
plt.show()
```

Negative Tweets Word Cloud