3.2-NLP-Final-Sentiment-Topic

December 5, 2023

0.0.1 Minh Vo

0.0.2 ADSP 32018 - Natural Language Processing and Cognitive Computing (Autumn 23)

0.0.3 FINAL PROJECT

0.0.4 Sentiment Over Time & By Topic Modeling

[8]:
```
# !pip install torch
# !pip install textblob
```

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import itertools
import re
import os, requests, sys
import nltk as nltk

from joblib import dump, load

import ktrain
from ktrain import text

import tensorflow as tf

import warnings
warnings.filterwarnings("ignore")

pd.set_option('display.max_rows', 100)
pd.set_option('display.max_columns', None)
pd.set_option('display.max_colwidth', 500)

from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer, TfidfTransformer
```

```
WARNING:tensorflow:From C:\Users\nhoxh\anaconda3\Lib\site-
packages\keras\src\losses.py:2976: The name
tf.losses.sparse_softmax_cross_entropy is deprecated. Please use
tf.compat.v1.losses.sparse_softmax_cross_entropy instead.
```

[2]:
```
import multiprocessing
from pandarallel import pandarallel

num_processors = multiprocessing.cpu_count()
print(f'Available CPUs: {num_processors}')
```

```
Available CPUs: 20
```

[3]:
```
pandarallel.initialize(nb_workers=num_processors-1, use_memory_fs=False)
workers=num_processors
```

```
INFO: Pandarallel will run on 19 workers.
INFO: Pandarallel will use standard multiprocessing data transfer (pipe) to
transfer data between the main process and workers.

WARNING: You are on Windows. If you detect any issue with pandarallel, be sure
you checked out the Troubleshooting page:
https://nalepae.github.io/pandarallel/troubleshooting/
```

News Data w/ Sentiment Loading

[37]:
```
%%time

news_df = pd.read_parquet('cleaned_sentiment_df.parquet', engine = 'pyarrow')

print(f'Data contains {news_df.shape[0]:,.0f} news articles')
```

```
Data contains 183,609 news articles
CPU times: user 10.2 s, sys: 7.18 s, total: 17.3 s
Wall time: 11.9 s
```

[38]:
```
news_df.head(3)
```

[38]:
```
                url  \
0  http://galusaustralis.com/2020/02/486473/legaltech-artificial-intelligence-
market-2019-technology-advancement-and-future-scope-casetext-inc-catalyst-
repository-systems-ebrevia/
1
http://spaceref.com/astronomy/observation-simulation-and-ai-join-forces-to-
reveal-a-clear-universe.html
2
http://usweekly.com/news/17/40964/Artificial-intelligence-yields-new-
```

```
antibiotic.html

         date language  \
0  2020-02-26       en
1  2021-07-05       en
2  2020-02-23       en

                                                                                title
\
0  LegalTech Artificial Intelligence Market 2019 Technology Advancement and
Future Scope - Casetext Inc., Catalyst Repository Systems, eBREVIA - Galus
Australis
1
Observation, Simulation, And AI Join Forces To Reveal A Clear Universe -
SpaceRef
2
Artificial intelligence yields new antibiotic - USweekly

                  text  \
0  LegalTech Artificial Intelligence Market 2019 Technology Advancement and
Future Scope - Casetext Inc., Catalyst Repository Systems, eBREVIA - Galus
Australis          \n\nGalus Australis\n\nBusinessGeneral
NewsHealthcareIndustryInternationalLifestyleSci-Tech\n\nWednesday, February 26
2020\n\nTrending\n\nNeedle Counters Market Comprehensive Study by Companies
Medline Industries, Boen Healthcare\nSkin Scrub Trays Market Comprehensive Study
by Companies Medline Industries, BD, Deroyal\nGlobal P…
1  \n\nObservation, Simulation, And AI Join Forces To Reveal A Clear Universe -
SpaceRef\n\n \nHome |\nNASA Watch\nSpaceRef Business\nAstrobiology
Web\nAdvertising\nAdd an Event\nSign up for our Daily Newsletter \n\n
\n\n\nInternational Space Station\nNASA Hack Space\nCalendar\nMissions\nSpace
Weather \n                            \n           \nObservation, Simulation, And
AI Join Forces To Reveal A Clear Universe\n\n\n             Press Release -
Source: NATIONAL INSTITUTES OF NATURAL SCIENCES…
2  \n\n\nArtificial intelligence yields new antibiotic - USweekly\nSunday, 23
February 2020\nSend search form\n \n\nTodays news\nWorld\nU.S.
National\nPolitics\nBusiness\nTechnology\nSports\nEntertainment\nBeauty &
Health\nLiving & Travel\nScience\nWeather\nOdd news\nShopping\n\n
\n\n\nArtificial intelligence yields new antibiotic\nAdded: 21.02.2020 21:19 |
35 views | 0 comments\nSource: www.slideshare.netUsing a machine-learning
algorithm, researchers have identified a powerful new antibiotic…

                                                  cleaned_title  \
0  legaltech artificial intelligence market technology advancement future scope
casetext inc catalyst repository systems ebrevia galus australis
1
observation simulation ai join forces reveal clear universe spaceref
2
```

```
artificial intelligence yields new antibiotic usweekly

          cleaned_text  \
0  legaltech artificial intelligence market technology advancement future scope
casetext inc catalyst repository systems ebrevia galus australis galus australis
wednesday february trending needle counters market comprehensive study companies
medline industries boen healthcare skin scrub trays market comprehensive study
companies medline industries bd deroyal global portable handheld electronic game
machine market outlook business insights apollo games sony aristocrat leisure
igt infectious dise…
1  observation simulation ai join forces reveal clear universe spaceref nasa
watch spaceref business astrobiology web advertising add event sign daily
international space station nasa hack space calendar missions space weather
observation simulation ai join forces reveal clear universe press release source
national institutes natural sciences posted july pm view using ai driven data
analysis peel back noise find actual shape universe credit institute statistical
mathematics japanese astronomers…
2  artificial intelligence yields new antibiotic usweekly sunday february send
form todays news world national politics business technology sports
entertainment beauty health living travel science weather odd news shopping
artificial intelligence yields new antibiotic added views source algorithm
researchers identified powerful new antibiotic compound laboratory tests drug
killed many worlds problematic bacteria including strains resistant known
antibiotics also cleared infections two different…

   text_word_count  \
0              666
1              587
2              136

lemmatized_title  \
0  legaltech artificial intelligence market technology advancement catalyst
repository system
1                         observation simulation ai join force reveal clear
universe spaceref
2                                      artificial intelligence yield new
antibiotic usweekly

       lemmatized_text  \
0  legaltech artificial intelligence market technology advancement catalyst
repository system galus trend needle counter market comprehensive study company
industry market comprehensive study company industry global portable handheld
electronic game machine market outlook business insight leisure infectious
disease testing use market comprehensive study company roche diagnostic
veterinary dental xray generator market comprehensive study company veterinary
market comprehensive study company hein…
1  observation simulation ai join force reveal clear universe watch business
```

```
astrobiology web advertising add event international space station hack space
calendar mission space weather observation simulation ai join force reveal clear
universe press release source national natural science post pm view use ai drive
datum analysis peel back noise find actual shape universe credit statistical
japanese astronomer develop new artificial intelligence ai technique remove
noise astronomical datum due …
2  artificial intelligence yield new send form today news national business
technology sport entertainment beauty health living travel science weather odd
news shopping artificial intelligence yield new antibiotic add view source
researcher identify powerful new antibiotic compound laboratory test drug kill
many world problematic bacteria include strain resistant known antibiotic also
clear infection different mouse modelsmore tag enter image code tag good buy
breast cancer cher climate change …

   Sentiments_NB  Sentiments_LR  Sentiments_SVM Sentiment_VADER  compound  \
0              0              0               0        Positive    0.9990
1              0              0               0        Positive    0.9937
2              0              0               0        Positive    0.1027

  Sentiment_TextBlob
0           Negative
1           Positive
2           Positive
```

[39]:
```
news_df.columns
```

[39]:
```
Index(['url', 'date', 'language', 'title', 'text', 'cleaned_title',
       'cleaned_text', 'text_word_count', 'lemmatized_title',
       'lemmatized_text', 'Sentiments_NB', 'Sentiments_LR', 'Sentiments_SVM',
       'Sentiment_VADER', 'compound', 'Sentiment_TextBlob'],
      dtype='object')
```

[40]:
```
news_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 183609 entries, 0 to 191486
Data columns (total 16 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   url                 183609 non-null  object
 1   date                183609 non-null  object
 2   language            183609 non-null  object
 3   title               183609 non-null  object
 4   text                183609 non-null  object
 5   cleaned_title       183609 non-null  object
 6   cleaned_text        183609 non-null  object
 7   text_word_count     183609 non-null  int64
```

```
 8   lemmatized_title    183609 non-null  object
 9   lemmatized_text     183609 non-null  object
 10  Sentiments_NB       183609 non-null  int64
 11  Sentiments_LR       183609 non-null  int64
 12  Sentiments_SVM      183609 non-null  int64
 13  Sentiment_VADER     183609 non-null  object
 14  compound            183609 non-null  float64
 15  Sentiment_TextBlob  183609 non-null  object
dtypes: float64(1), int64(4), object(11)
memory usage: 23.8+ MB
```

[41]:
```
news_df = news_df[['date', 'title', 'text', 'cleaned_title','cleaned_text',␣
↪'text_word_count',
                   'lemmatized_title','lemmatized_text','Sentiment_TextBlob']]
```

Sentiment Over Time

[42]:
```
news_df['Sentiment_TextBlob'].value_counts()
```

[42]:
```
Sentiment_TextBlob
Positive    170688
Negative     12881
Neutral         40
Name: count, dtype: int64
```

[43]:
```
sentiment_time = news_df[['Sentiment_TextBlob', 'date']]
sentiment_time['date'] = pd.to_datetime(sentiment_time['date'])
sentiment_time.head()
```

[43]:

	Sentiment_TextBlob	date
0	Negative	2020-02-26
1	Positive	2021-07-05
2	Positive	2020-02-23
3	Positive	2021-03-26
4	Positive	2023-07-21

[44]:
```
sentiment_time['year-month'] = sentiment_time['date'].dt.strftime('%Y-%m')
sentiment_time.head()
```

[44]:

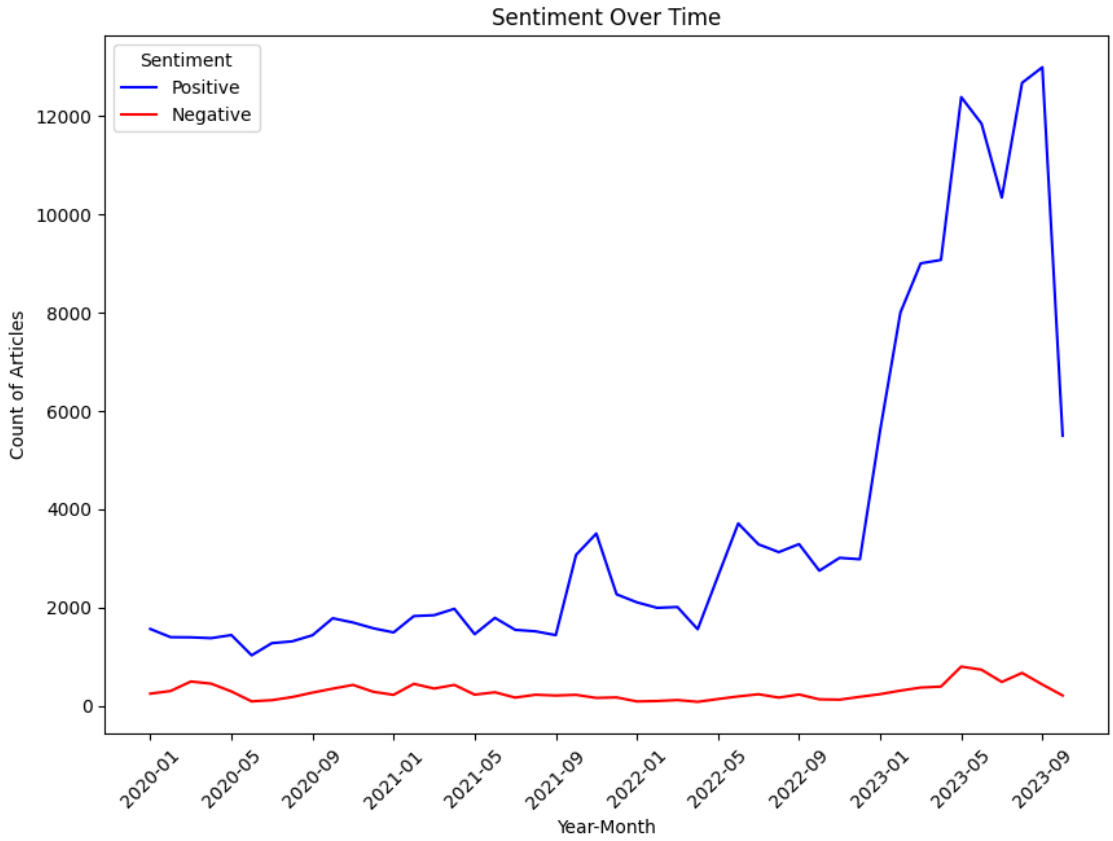
	Sentiment_TextBlob	date	year-month
0	Negative	2020-02-26	2020-02
1	Positive	2021-07-05	2021-07
2	Positive	2020-02-23	2020-02
3	Positive	2021-03-26	2021-03
4	Positive	2023-07-21	2023-07

[45]:
```
sentiment_time = sentiment_time.groupby(['year-month', 'Sentiment_TextBlob']).
 ↪size().reset_index(name='count')
sentiment_time.head()
```

[45]:

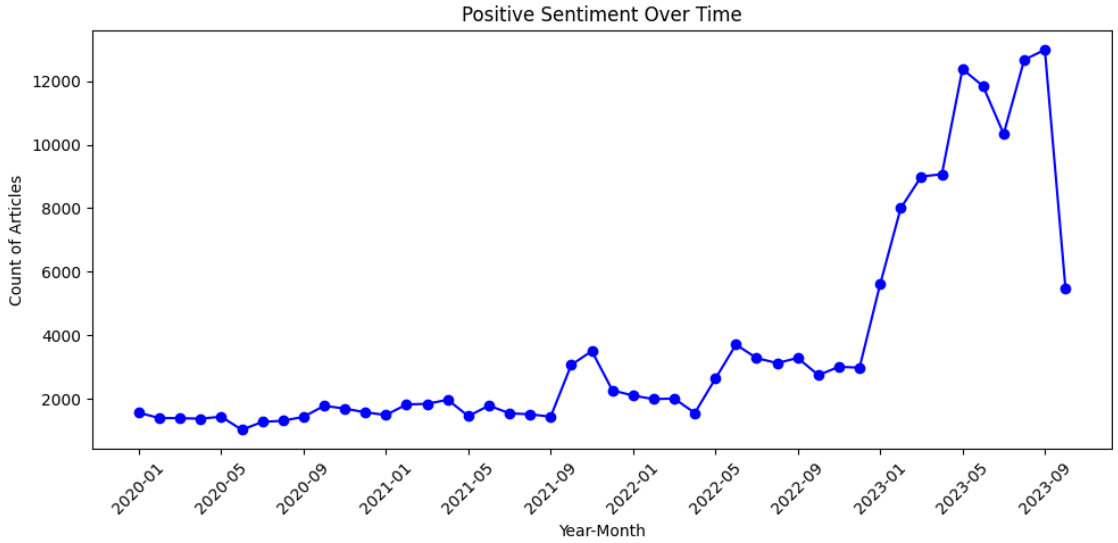
	year-month	Sentiment_TextBlob	count
0	2020-01	Negative	244
1	2020-01	Positive	1560
2	2020-02	Negative	298
3	2020-02	Neutral	1
4	2020-02	Positive	1393

[48]:
```
# Pivot the DataFrame to make it suitable for line plotting
sentiment_pivot = sentiment_time.pivot(index='year-month',␣
 ↪columns='Sentiment_TextBlob', values='count').fillna(0)

# Plotting
plt.figure(figsize=(10,7))
colors = {'Positive': 'blue', 'Negative': 'red'}

# Plot each sentiment with its respective color
for sentiment, color in colors.items():
    if sentiment in sentiment_pivot.columns:
        plt.plot(sentiment_pivot.index, sentiment_pivot[sentiment],␣
 ↪label=sentiment, color=color)

plt.title('Sentiment Over Time')
plt.xlabel('Year-Month')
plt.ylabel('Count of Articles')
plt.legend(title='Sentiment')
plt.xticks(rotation=45)
plt.gca().xaxis.set_major_locator(plt.MaxNLocator(15))
plt.show()
```

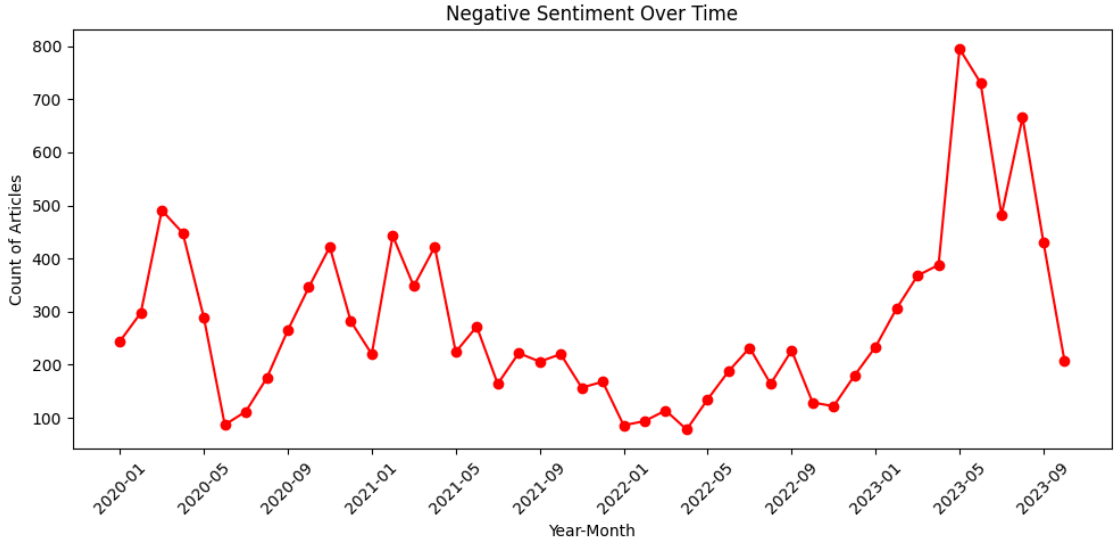


[49]:
```
# Now let's try plotting again, this time ensuring we have the 'Positive' and␣
 ↪'Negative' columns available.
def plot_sentiment_over_time(sentiment, color, title):
    plt.figure(figsize=(10,5))
    if sentiment in sentiment_pivot.columns:
        plt.plot(sentiment_pivot.index, sentiment_pivot[sentiment],␣
 ↪label=sentiment, color=color, marker='o')
        plt.title(title)
        plt.xlabel('Year-Month')
        plt.ylabel('Count of Articles')
        plt.xticks(rotation=45)
        plt.gca().xaxis.set_major_locator(plt.MaxNLocator(15))
        plt.tight_layout()  # Adjust layout to prevent clipping of tick-labels
        plt.show()
    else:
        print(f"The sentiment '{sentiment}' is not found in the DataFrame.")
```

[50]:
```
# Plot for Positive sentiment
plot_sentiment_over_time('Positive', 'blue', 'Positive Sentiment Over Time')
```



[51]:
```
# Plot for Negative sentiment
plot_sentiment_over_time('Negative', 'red', 'Negative Sentiment Over Time')
```



Topic Modeling by Sentiment - Zero-shot Learning

Check for GPU presence

[5]:
```
#Verify we got CPU + GPU or only CPU
tf.config.list_physical_devices()
```

[5]:
```
[PhysicalDevice(name='/physical_device:CPU:0', device_type='CPU')]
```

[6]:
```
!nvidia-smi
```

```
Wed Nov 29 14:11:48 2023
+---------------------------------------------------------------------------
--------+
| NVIDIA-SMI 532.09                 Driver Version: 532.09       CUDA Version:
12.1     |
|-----------------------------------------+----------------------+----------
--------+
| GPU  Name                      TCC/WDDM | Bus-Id        Disp.A | Volatile
Uncorr. ECC |
| Fan  Temp  Perf            Pwr:Usage/Cap|         Memory-Usage | GPU-Util
Compute M. |
|                                         |                      |
MIG M. |
|=========================================+======================+==========
========|
|   0  NVIDIA GeForce RTX 3050 T…  WDDM | 00000000:01:00.0 Off |
N/A |
| N/A   50C    P3               10W /  N/A|      0MiB /  4096MiB |      0%
Default |
|                                         |                      |
N/A |
+-----------------------------------------+----------------------+----------
--------+

+---------------------------------------------------------------------------
--------+
| Processes:
|
|  GPU   GI   CI        PID   Type   Process name                            GPU
Memory |
|        ID   ID
Usage      |
|=========================================================================
========|
|  No running processes found
|
+---------------------------------------------------------------------------
--------+
```

[7]:
```
tf.__version__
```

[7]:
```
'2.15.0'
```

[39]:
```
zsl = ktrain.text.ZeroShotClassifier()
```

[16]:
```
topic_labels=['healthcare', 'news platform', 'customer experience',␣
 ↪'education', 'finance', 'global market', 'sports',
              'media', 'cybersecurity', 'robotic', 'chatgpt', 'image',␣
 ↪'automation', 'trade', 'blockchain']

# Function to predict topics for a batch of texts
def predict_topics(texts):
    predictions = zsl.predict(texts,
                        labels=topic_labels,
                        include_labels=False,
                        nli_template="The topic of this news text is {}.",
                        batch_size=64)
    return predictions
```

Positive Articles

[16]:
```
positive_news_df = news_df[news_df['Sentiment_TextBlob'] == 'Positive']
# positive_news_df = positive_news_df[['date', 'title', 'text', 'cleaned_title',
#         'cleaned_text', 'text_word_count', 'lemmatized_title',
#         'lemmatized_text','Sentiment_TextBlob']]
positive_news_df.shape
```

[16]:
```
(170688, 9)
```

[55]:
```
# Save this positive news df
positive_news_df.rename(columns={'Sentiment_TextBlob': 'Sentiment'},␣
 ↪inplace=True)

positive_news_df.to_parquet('positive_news_df.parquet')

bucket_name = 'minhvo-nlp'
file_path = 'positive_news_df.parquet'

storage_client = storage.Client()
bucket = storage_client.get_bucket(bucket_name)
blob = bucket.blob(file_path)
blob.upload_from_filename(file_path)
```

[4]:
```
positive_news_df = pd.read_parquet('positive_news_df.parquet', engine =␣
 ↪'pyarrow')
positive_news_df = positive_news_df.reset_index()
```

[5]:
```
positive_news_df.sample(1)
```

[5]:
```
       index        date  \
25656  28794  2021-02-20

                                                                  title  \
25656  Google fires AI manager who protested her peer’s departure - WAVY.com

                      text  \
25656  \n \nGoogle fires AI manager who protested her peer’s departure -
WAVY.com\nSkip to content\n\n\nWAVY.com\nNorfolk\n\n30°\nSponsored By\n\n
\nToggle Menu\n\n\nOpen Navigation\nClose Navigation\n\nSearch\nPrimary
Menu\n\nNews\nLocal News\nCoronavirus\nCOVID-19
Vaccine\nInvestigative\nCrime\nMilitary\nVirginia\nNorth
Carolina\nNational\nNewsfeed Now\nD. C. Bureau\nPolitics\nPass or
Fail\nStrange\n10 On Your Side\nRemarkable Women\nBlack History
Month\nCourageous Conversations\nCoats for Famil…

                                                cleaned_title  \
25656  google fires ai manager protested peers departure wavycom

              cleaned_text  \
25656  google fires ai manager protested peers departure wavycom content wavycom
norfolk toggle open navigation close navigation primary news local news
coronavirus covid vaccine investigative crime military virginia north carolina
national newsfeed c bureau politics pass fail strange side remarkable women
black history month courageous conversations coats families investigation
patients v perwaiz vaccinate virginia questions answered top stories suffolk
police searching endangered missing teen rep…

       text_word_count                                lemmatized_title  \
25656              810  fire ai manager protest peer departure wavycom

           lemmatized_text  \
25656  fire ai manager protest peer departure open navigation close navigation
primary news local news coronavirus investigative crime pass fail strange side
remarkable woman black history month courageous conversation coat family
investigation patient question answer top story suffolk police search endangered
miss teen raise winter weather crisis close reversible lane night video
perquiman sheriff office warn scammer impersonate deputy video live stream live
break news archive weather alert closin…

      Sentiment
25656  Positive
```

[9]:
```
positive_news_df.columns
```

[9]:
```
Index(['index', 'date', 'title', 'text', 'cleaned_title', 'cleaned_text',
       'text_word_count', 'lemmatized_title', 'lemmatized_text', 'Sentiment'],
```

```
      dtype='object')
```

Retrieve a sample data of 1000 rows

[21]:
```
sample_pos = positive_news_df.sample(n=1000, random_state=42)
sample_pos.shape
```

[21]:
```
(1000, 10)
```

[54]:
```
%%time

sample_pos_text = sample_pos['cleaned_text'].to_list()
# Apply the function to  positive news texts
topic_positive = predict_topics(sample_pos_text)
topic_positive_df = pd.DataFrame(topic_positive, columns=topic_labels)
```

```
CPU times: user 1d 19h 34min 55s, sys: 13h 4min 20s, total: 2d 8h 39min 15s
Wall time: 1h 46min 22s
```

[55]:
```
print(topic_positive_df.shape)
topic_positive_df.head()
```

```
(1000, 15)
```

[55]:

	healthcare	news platform	customer experience	education	finance	global market	sports	media	cybersecurity	robotic	chatgpt	image	automation	trade	blockchain
0	0.070244	0.934708	0.791054	0.110341	0.028165	0.522165	0.010822	0.940283	0.242465	0.101374	0.625605	0.998557	0.582430	0.172931	0.233095
1	0.019375	0.952176	0.556920	0.092113	0.998849	0.992299	0.008210	0.916207	0.173164	0.139236	0.864229	0.644893	0.142193	0.975875	0.050060
2	0.109741	0.902125	0.973460	0.179841	0.483347	0.805206	0.185778	0.990876	0.946868	0.315302	0.704714	0.883268	0.980408	0.668269	0.094272
3	0.697724	0.885077	0.752874	0.481156	0.222054	0.255113	0.045973	0.989007	0.562478	0.017129	0.996047	0.136368	0.489922	0.257482	0.061999
4	0.554344	0.856028	0.758342	0.133531	0.067148	0.133585	0.624725	0.977159	0.326418	0.054735	0.994289	0.813283	0.494408	0.258292	0.031913

[22]:
```
# Save the result
# topic_positive_df.to_json('topic_positive_df.json', orient='records',␣
 ↪lines=True)
```

```
topic_positive_df = pd.read_json('topic_positive_df.json', orient='records',␣
↪lines=True)
```

[36]:
```
positive_news = topic_positive_df.join(positive_news_df, how='inner')
positive_news = positive_news[['cleaned_text','Sentiment'] + topic_labels]

positive_news.columns = positive_news.columns.get_level_values(0)
print(positive_news.shape)
positive_news.head(2)
```

```
(1000, 17)
```

[36]:

	cleaned_text	Sentiment	healthcare	news platform	customer experience	education	finance	global market	sports	media	cybersecurity	robotic	chatgpt	image	automation	trade	blockchain
0	observation simulation ai join forces reveal clear universe spaceref nasa watch spaceref business astrobiology web advertising add event sign daily international space station nasa hack space calendar missions space weather observation simulation ai join forces reveal clear universe press release source national institutes natural sciences posted july pm view using ai driven data analysis peel back noise find actual shape universe credit institute statistical mathematics japanese astronomers…	Positive	0.070244	0.934708	0.791054	0.110341	0.028165	0.522165	0.010822	0.940283	0.242465	0.101374	0.625605	0.998557	0.582430	0.172931	0.233095
1	artificial intelligence yields new antibiotic usweekly sunday february send form todays news world national politics business technology sports entertainment beauty health living travel science weather odd news shopping artificial intelligence yields new antibiotic added views source algorithm researchers identified powerful new antibiotic compound laboratory tests drug killed many worlds problematic bacteria including strains resistant known antibiotics also cleared infections two different…	Positive	0.019375	0.952176	0.556920	0.092113	0.998849	0.992299	0.008210	0.916207	0.173164	0.139236	0.864229	0.644893	0.142193	0.975875	0.050060

[28]:
```
# Save the result
positive_news.to_json('sample_pos_news_text.json', orient='records', lines=True)
```

Full Data (on news TITLE)

[64]:
```
pos_title_list = positive_news_df['cleaned_title'].tolist()
```

[ ]:
```
%%time

# Apply the function to  positive news TITLE
topic_positive_title = predict_topics(pos_title_list)
topic_positive_title_df = pd.DataFrame(topic_positive_title,␣
 ↪columns=topic_labels)
```

[ ]:
```
print(topic_positive_title_df.shape)
topic_positive_title_df.head()
```

[ ]:
```
# Save the result
topic_positive_title_df.to_json('topic_positive_title_df.json',␣
 ↪orient='records', lines=True)
```

[ ]:
```
positive_news2 = topic_positive_title_df.join(positive_news_df, how='inner')
positive_news2 = positive_news2[['cleaned_title','cleaned_text','Sentiment'] +␣
 ↪topic_labels]

print(positive_news2.shape)
positive_news2.head(2)
```

[ ]:
```
# Save the result
positive_news2.to_json('pos_news_title.json', orient='records', lines=True)
```

Negative Articles

[21]:
```
negative_news_df = news_df[news_df['Sentiment_TextBlob'] == 'Negative']
negative_news_df.shape
```

[21]:
```
(12881, 9)
```

[56]:
```
# Save this negative news df
negative_news_df.rename(columns={'Sentiment_TextBlob': 'Sentiment'},␣
 ↪inplace=True)

negative_news_df.to_parquet('negative_news_df.parquet')

bucket_name = 'minhvo-nlp'
file_path = 'negative_news_df.parquet'

storage_client = storage.Client()
bucket = storage_client.get_bucket(bucket_name)
blob = bucket.blob(file_path)
blob.upload_from_filename(file_path)
```

[10]:
```
negative_news_df = pd.read_parquet('negative_news_df.parquet', engine =␣
 ↪'pyarrow')
```

```
negative_news_df = negative_news_df.reset_index()
```

[11]:
```
negative_news_df.sample(1)
```

[11]:
```
       index        date  \
7527  112692  2023-10-16

                                                                  title  \
7527  Remark Holdings is recognized as a representative vendor for Computer
Vision Perceptive Systems by Gartner AI Analyst Erick Brethenoux in his research
report, What is Artificial Intelligence? Ignore the hype; Here's where to start

                      text  \
7527  Remark Holdings is recognized as a representative vendor for Computer
Vision Perceptive Systems by Gartner AI Analyst Erick Brethenoux in his research
report, What is Artificial Intelligence? Ignore the hype; Here's where to
start\n\nSkip to
contentFacebookInstagramTwitterLinkedInNewsLiveVideoWAFB+WeatherHurricane
CenterI-TEAMBrave CaveSportsElectionsWAFB Channel 9NewsWAFB 70th
AnniversaryCrime9News NowTrafficWatching Your WalletElectionsVideoI-TEAMBrave
Cave9News AlertInvestigate TVWeatherF…

 cleaned_title  \
7527  remark holdings recognized vendor computer vision perceptive systems
gartner ai analyst erick brethenoux research report artificial intelligence
ignore hype heres start

             cleaned_text  \
7527  remark holdings recognized vendor computer vision perceptive systems
gartner ai analyst erick brethenoux research report artificial intelligence
ignore hype heres start channel newswafb th cavenews alert radar networktitan
skyhurricane centerth seasoncolor awardshigh schoolplayer talkgame timestats
uplouisiana weekendsubmit photos videoask attorneyask expertheart upabout
uscontact usadvertise uswafb dc bureaucircle country music holdings recognized
vendor computer vision perceptive systems g…

      text_word_count  \
7527              601

                     lemmatized_title  \
7527  holding recognize vendor computer vision analyst research report
artificial intelligence ignore hype here start

          lemmatized_text  \
7527  holding recognize vendor computer vision analyst research report
artificial intelligence ignore hype here start newswafb cavenew alert
seasoncolor awardshigh schoolplayer talkgame timestat weekendsubmit photo
```

```
uscontact usadvertise uswafb country music holding recognize vendor computer
vision analyst research report artificial intelligence ignore hype here oct
cdtupdated hour leading provider artificial intelligence aipowere video analytic
solution recognize vendor computer vision analyst res…

     Sentiment
7527  Negative
```

Retrieve a sample data of 1000 rows

[33]:
```
sample_neg = negative_news_df.sample(n=1000, random_state=42)
sample_neg_text = sample_neg['cleaned_text'].tolist()
```

[ ]:
```
%%time

# Apply the function
topic_negative = predict_topics(sample_neg_text)
topic_negative_df = pd.DataFrame(topic_negative, columns=topic_labels)
```

[ ]:
```
print(topic_negative_df.shape)
topic_negative_df.head()
```

[32]:
```
# Save the result
# topic_negative_df.to_json('topic_negative_df.json', orient='records',␣
 ↪lines=True)

topic_negative_df = pd.read_json('topic_negative_df.json', orient='records',␣
 ↪lines=True)
topic_negative_df.head()
```

[32]:

	healthcare	news platform	customer experience	education	finance	global market	sports	media	cybersecurity	robotic	chatgpt	image	automation	trade	blockchain
0	0.178577	0.786087	0.737141	0.354214	0.169157	0.979307	0.038063	0.919971	0.472614	0.279459	0.726452	0.760226	0.858769	0.407282	0.121111
1	0.994953	0.797421	0.037159	0.006153	0.003151	0.113485	0.001623	0.994452	0.035213	0.216393	0.014611	0.010333	0.922993	0.004440	0.001964
2	0.287461	0.977536	0.824716	0.124460	0.854959	0.991544	0.012048	0.985268	0.874412	0.151401	0.871122				
3	0.106281	0.968851	0.917921	0.098798	0.460349	0.953076	0.515453	0.977528	0.759604	0.126416	0.808916				
4	0.790292	0.593470	0.724533	0.158518	0.003653	0.138400	0.436984	0.970279	0.070862	0.033752	0.969495				

```
2  0.602833    0.946153  0.582133    0.060945
3  0.909420    0.980746  0.473662    0.273704
4  0.487801    0.836421  0.014547    0.005256
```

[34]:
```
negative_news = topic_negative_df.join(negative_news_df, how='inner')
negative_news = negative_news[['cleaned_text','Sentiment'] + topic_labels]

negative_news.columns = negative_news.columns.get_level_values(0)
print(negative_news.shape)
negative_news.head(2)
```

```
(1000, 17)
```

[34]:
```
           cleaned_text  \
0  legaltech artificial intelligence market technology advancement future scope
casetext inc catalyst repository systems ebrevia galus australis galus australis
wednesday february trending needle counters market comprehensive study companies
medline industries boen healthcare skin scrub trays market comprehensive study
companies medline industries bd deroyal global portable handheld electronic game
machine market outlook business insights apollo games sony aristocrat leisure
igt infectious dise…
1  artificial intelligence behavioral mental health care market witness
astonishing growth focusing leading players advancedmd cerner core solutions
credible behavioral health w market news reports content sunday june contact w
market news reports rd market reports analytics news market reports industry
analytics industry reports market research business opportunity emerging trends
growth prospects intelligence behavioral mental health care market witness
astonishing growth focusing leading pla…

  Sentiment  healthcare  news platform  customer experience  education  \
0  Negative    0.178577       0.786087             0.737141   0.354214
1  Negative    0.994953       0.797421             0.037159   0.006153

    finance  global market    sports     media  cybersecurity   robotic  \
0  0.169157       0.979307  0.038063  0.919971       0.472614  0.279459
1  0.003151       0.113485  0.001623  0.994452       0.035213  0.216393

    chatgpt     image  automation     trade  blockchain
0  0.726452  0.760226    0.858769  0.407282    0.121111
1  0.014611  0.010333    0.922993  0.004440    0.001964
```

[ ]:
```
predictions_df['topic'] = predictions_df.apply(lambda row: row.idxmax(), axis=1)
```

[37]:
```
# Save the result
negative_news.to_json('sample_neg_news_text.json', orient='records', lines=True)
```

Full Data (on news TITLE)

[42]:
```
%%time

neg_title_list = negative_news_df['cleaned_title'].to_list()

# Apply the function to  negative news TITLE
topic_negative_title = predict_topics(neg_title_list)
topic_negative_title_df = pd.DataFrame(topic_negative_title,␣
 ↪columns=topic_labels)
```

[43]:
```
print(topic_negative_title_df.shape)
topic_negative_title_df.head()
```

[ ]:
```
# Save the result
# topic_negative_title_df.to_json('topic_negative_title_df.json',␣
 ↪orient='records', lines=True)
```

[ ]:
```
negative_news2 = topic_negative_title_df.join(negative_news_df, how='inner')
negative_news2 = negative_news2[['cleaned_title','cleaned_text','Sentiment'] +␣
 ↪topic_labels]

negative_news2.columns = negative_news2.columns.get_level_values(0)
print(negative_news2.shape)
negative_news2.head(2)
```

Topic Modeling by BERTopic

[7]:
```
# !pip install bertopic
import ast
import os
os.environ["CUDA_VISIBLE_DEVICES"] = ""
os.environ["TOKENIZERS_PARALLELISM"] = "false"

import bertopic
from bertopic import BERTopic
```

Positive Articles

[8]:
```
positive_news_df.sample(1)
```

[8]:
```
       index        date  \
59368  66633  2023-02-25

                                                                 title  \
59368     Brain Training on the rise as AI Dominates the Workplace

                       text  \
59368  \n  Brain Training on the rise as AI Dominates the Workplace\n  \n
About\n        \n            About EIN Presswire\n          \n            How We
```

```
Are Different. Better\n          \n            How It Works\n          \n
Testimonials\n          \n            Contact\n          \n            EIN
Presswire in the News\n          \n\n          Pricing\n        \n
Distribution\n        \n            Distribution Overview\n          \n
Media Database\n    …

                                    cleaned_title  \
59368  brain training rise ai dominates workplace

              cleaned_text  \
59368  brain training rise ai dominates workplace ein presswire different better
works testimonials contact ein presswire news pricing distribution distribution
overview media database major news sites tv radio stations international
newswires newswires industry newswires country newswires state mobile apps
newsplugin live feed sample distribution report press releases featured industry
country state archive newswires international newswires newswires industry
agriculture airline automotive banking…

       text_word_count                  lemmatized_title  \
59368              954  brain training rise ai workplace

           lemmatized_text  \
59368  brain training rise ai presswire different well work testimonial contact
presswire news pricing distribution distribution overview medium database major
news site tv radio station newswire mobile app newsplugin live feed sample
distribution report press release feature industry country state archive
newswire agriculture airline automotive banking book publish business casino
chemical company conference trade show construction consumer cosmetic education
electronic emergency service energy en…

      Sentiment
59368  Positive
```

[23]:
```
%%time

mod_BERT_pos = BERTopic(calculate_probabilities=True, verbose=True,␣
 ↪min_topic_size=50)
topics_pos, probabilities_pos = mod_BERT_pos.
 ↪fit_transform(positive_news_df['cleaned_text'].tolist())
```

```
2023-12-04 16:00:17,245 - BERTopic - Embedding - Transforming documents to
embeddings.

Batches:   0%|          | 0/5334 [00:00<?, ?it/s]

2023-12-04 18:12:55,621 - BERTopic - Embedding - Completed
2023-12-04 18:12:55,625 - BERTopic - Dimensionality - Fitting the dimensionality
reduction algorithm
```

```
2023-12-04 18:17:59,655 - BERTopic - Dimensionality - Completed
2023-12-04 18:17:59,668 - BERTopic - Cluster - Start clustering the reduced
embeddings
2023-12-04 21:36:50,159 - BERTopic - Cluster - Completed
2023-12-04 21:36:50,576 - BERTopic - Representation - Extracting topics from
clusters using representation models.
2023-12-04 21:41:13,064 - BERTopic - Representation - Completed

CPU times: total: 8h 25min 45s
Wall time: 5h 42min 27s
```

[24]:
```
print("Number of topics:", mod_BERT_pos.get_topic_freq().shape[0])
```

```
Number of topics: 643
```

[25]:
```
positive_news_df['BERTopic']=mod_BERT_pos.topics_
positive_news_df['BERTopic_words'] = positive_news_df['BERTopic'].apply(lambda␣
↪x: mod_BERT_pos.get_topic(x))
```

[26]:
```
print(positive_news_df.shape)
positive_news_df.head(3)
```

```
(170688, 12)
```

[26]:
```
   index        date  \
0      1  2021-07-05
1      2  2020-02-23
2      3  2021-03-26

                                          title  \
0                                    Observation, Simulation, And AI Join
Forces To Reveal A Clear Universe - SpaceRef
1                                                                  Artificial
intelligence yields new antibiotic - USweekly
2  Forget ML, AI and Industry 4.0 - obsolescence should be your focus - 26
February 2021 - Test & Rework Solutions - Dataweek

                  text  \
0  \n\nObservation, Simulation, And AI Join Forces To Reveal A Clear Universe -
SpaceRef\n\n \nHome |\nNASA Watch\nSpaceRef Business\nAstrobiology
Web\nAdvertising\nAdd an Event\nSign up for our Daily Newsletter \n\n
\n\n\nInternational Space Station\nNASA Hack Space\nCalendar\nMissions\nSpace
Weather \n                            \n          \nObservation, Simulation, And
AI Join Forces To Reveal A Clear Universe\n\n\n             Press Release -
Source: NATIONAL INSTITUTES OF NATURAL SCIENCES…
1  \n\n\nArtificial intelligence yields new antibiotic - USweekly\nSunday, 23
February 2020\nSend search form\n \n\nTodays news\nWorld\nU.S.
National\nPolitics\nBusiness\nTechnology\nSports\nEntertainment\nBeauty &
```

```
Health\nLiving & Travel\nScience\nWeather\nOdd news\nShopping\n\n
\n\n\nArtificial intelligence yields new antibiotic\nAdded: 21.02.2020 21:19 |
35 views | 0 comments\nSource: www.slideshare.netUsing a machine-learning
algorithm, researchers have identified a powerful new antibiotic…
2  \n\nForget ML, AI and Industry 4.0 - obsolescence should be your focus - 26
February 2021 - Test & Rework Solutions - Dataweek\nHome\nAbout us\nBack issues
/ E-book / PDF\nEMP Handbook\nSubscribe\nAdvertise\n\nCategories\n\n  Editor's
Choice\n  Multimedia, Videos\n  Analogue, Mixed Signal, LSI\n  Circuit & System
Protection\n  Computer/Embedded Technology\n  Design Automation\n  DSP, Micros &
Memory\n  Electronics Technology\n  Enclosures, Racks, Cabinets & Panel
Products\n  Events\n  Interc…

cleaned_title  \
0              observation simulation ai join forces reveal clear universe
spaceref
1                            artificial intelligence yields new antibiotic
usweekly
2  forget ml ai industry obsolescence focus february test rework solutions
dataweek

          cleaned_text  \
0  observation simulation ai join forces reveal clear universe spaceref nasa
watch spaceref business astrobiology web advertising add event sign daily
international space station nasa hack space calendar missions space weather
observation simulation ai join forces reveal clear universe press release source
national institutes natural sciences posted july pm view using ai driven data
analysis peel back noise find actual shape universe credit institute statistical
mathematics japanese astronomers…
1  artificial intelligence yields new antibiotic usweekly sunday february send
form todays news world national politics business technology sports
entertainment beauty health living travel science weather odd news shopping
artificial intelligence yields new antibiotic added views source algorithm
researchers identified powerful new antibiotic compound laboratory tests drug
killed many worlds problematic bacteria including strains resistant known
antibiotics also cleared infections two different…
2  forget ml ai industry obsolescence focus february test rework solutions
dataweek back issues ebook emp handbook advertise categories editors choice
multimedia videos analogue mixed signal lsi circuit system protection technology
design automation dsp micros memory electronics technology enclosures racks
cabinets panel products events manufacturing production technology hardware
services news passive components power electronics power management programmable
logic smart automation switches re…

   text_word_count  \
0              587
1              136
2             1201
```

```
                                                    lemmatized_title  \
0  observation simulation ai join force reveal clear universe spaceref
1                artificial intelligence yield new antibiotic usweekly
2         forget ai industry obsolescence focus test solution dataweek

       lemmatized_text  \
0  observation simulation ai join force reveal clear universe watch business
astrobiology web advertising add event international space station hack space
calendar mission space weather observation simulation ai join force reveal clear
universe press release source national natural science post pm view use ai drive
datum analysis peel back noise find actual shape universe credit statistical
japanese astronomer develop new artificial intelligence ai technique remove
noise astronomical datum due …
1  artificial intelligence yield new send form today news national business
technology sport entertainment beauty health living travel science weather odd
news shopping artificial intelligence yield new antibiotic add view source
researcher identify powerful new antibiotic compound laboratory test drug kill
many world problematic bacteria include strain resistant known antibiotic also
clear infection different mouse modelsmore tag enter image code tag good buy
breast cancer cher climate change …
2  forget ai industry obsolescence focus test solution dataweek back issue ebook
handbook advertise category editor choice multimedia video analogue mixed signal
lsi circuit system protection technology design automation dsp micro memory
electronic technology enclosure rack cabinet panel product event manufacture
production technology hardware service news passive component power electronic
programmable logic smart automation switch relay telecom test category editor
choice multimedia video ana…

  Sentiment  BERTopic  \
0  Positive        37
1  Positive        -1
2  Positive        -1

                                                    BERTopic_words
0  [(nasa, 0.01537953437437768), (le, 0.013152422612171255), (space,
0.012979653683907038), (earth, 0.01136794940989286), (solar,
0.010636634206885376), (exoplanets, 0.01014066219141315), (satellite,
0.009847465108901582), (planets, 0.00843508836619673), (mars,
0.008143535760525532), (nasas, 0.0073258911898772235)]
1  [(gray, 0.0016537138818333377), (ai, 0.001597423445033531), (group,
0.0014650405127839617), (data, 0.001421173344257893), (new,
0.00137364641536202), (media, 0.0013709240655670404), (press,
0.0013682598692371963), (content, 0.0013144986514689585), (platform,
0.0013051026536570573), (news, 0.0012859285868718762)]
2  [(gray, 0.0016537138818333377), (ai, 0.001597423445033531), (group,
0.0014650405127839617), (data, 0.001421173344257893), (new,
```

```
0.00137364641536202), (media, 0.0013709240655670404), (press,
0.0013682598692371963), (content, 0.0013144986514689585), (platform,
0.0013051026536570573), (news, 0.0012859285868718762)]
```

[27]:
```
# Save the model
mod_BERT_pos.save("BERTopic_pos_model", serialization="pickle")
```

```
2023-12-04 21:45:50,111 - BERTopic - WARNING: When you use `pickle` to save/load
a BERTopic model,please make sure that the environments in which you saveand
load the model are **exactly** the same. The version of BERTopic,its
dependencies, and python need to remain the same.
```

[53]:
```
# Load from file
mod_BERT_pos = BERTopic.load("BERTopic_pos_model")
print("Number of topics:", mod_BERT_pos.get_topic_freq().shape[0])
```

```
Number of topics: 643
```

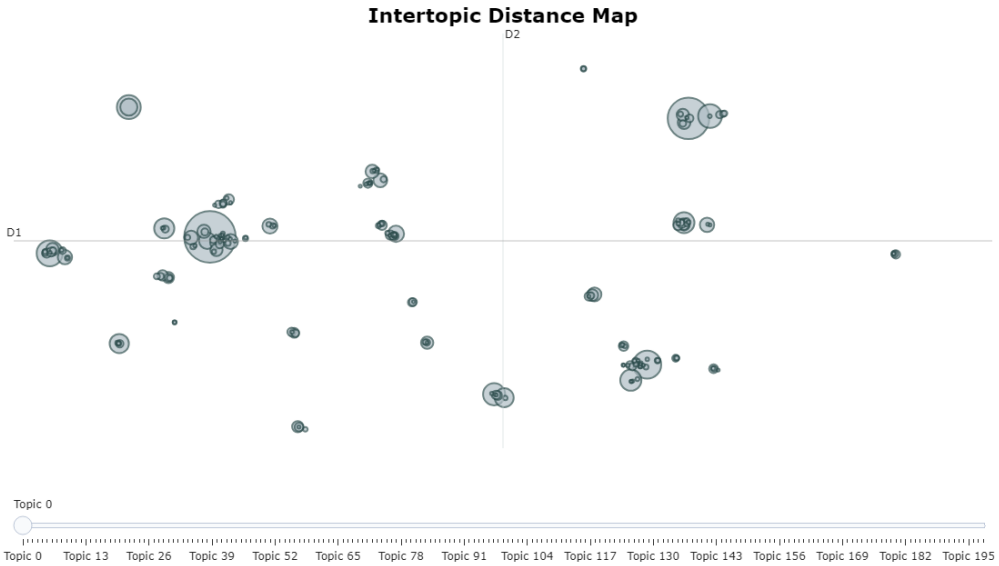
[54]:
```
mod_BERT_pos2=mod_BERT_pos.reduce_topics(positive_news_df['cleaned_text'],␣
↪nr_topics=200)
```

```
2023-12-05 00:17:30,101 - BERTopic - Topic reduction - Reducing number of topics
2023-12-05 00:20:26,239 - BERTopic - Topic reduction - Reduced number of topics
from 643 to 200
```

[55]:
```
print("Number of topics:", mod_BERT_pos2.get_topic_freq().shape[0])
```

```
Number of topics: 200
```

[56]:
```
mod_BERT_pos2.visualize_topics()
```



[22]:
```
positive_news_df.to_csv('positive_BERTopic_df.csv',index=False)
```

Negative Articles

[43]:
```
%%time

mod_BERT_neg = BERTopic(calculate_probabilities=True, verbose=True,␣
 ↪min_topic_size=50)
topics_neg, probabilities_neg = mod_BERT_neg.
 ↪fit_transform(negative_news_df['cleaned_text'].tolist())
```

```
2023-11-29 19:19:51,132 - BERTopic - Embedding - Transforming documents to
embeddings.
```

```
Batches:   0%|          | 0/403 [00:00<?, ?it/s]
```

```
2023-11-29 19:28:08,981 - BERTopic - Embedding - Completed
2023-11-29 19:28:08,984 - BERTopic - Dimensionality - Fitting the dimensionality
reduction algorithm
2023-11-29 19:28:20,170 - BERTopic - Dimensionality - Completed
2023-11-29 19:28:20,175 - BERTopic - Cluster - Start clustering the reduced
embeddings
2023-11-29 19:28:23,167 - BERTopic - Cluster - Completed
2023-11-29 19:28:23,204 - BERTopic - Representation - Extracting topics from
clusters using representation models.
2023-11-29 19:28:37,453 - BERTopic - Representation - Completed
```

```
CPU times: total: 9min 16s
Wall time: 8min 50s
```

[44]:
```
mod_BERT_neg.get_topic_info().head(10)
```

[44]:

	Topic	Count	Name	Representation
0	-1	4196	-1_ai_news_artificial_intelligence	[ai, news, artificial, intelligence, said, market,
1	0	1680	0_market_intelligence_artificial_report	
2	1	712	1_healthcare_market_medicine_intelligence	
3	2	602	2_market_global_analysis_intelligence	
4	3	355	3_etf_stock_robotics_llc	
5	4	298	4_radio_schedule_npr_public	
6	5	241	5_security_cyber_market_cybersecurity	
7	6	227	6_chipsets_chip_chips_chipset	
8	7	227	7_news_weather_sports_local	
9	8	208	8_funds_distribution_fund_nav	

```
technology, new, data, use]
1    [market, intelligence, artificial, report, analysis, global, growth,
industry, forecast, research]
2   [healthcare, market, medicine, intelligence, artificial, medical, analysis,
report, growth, global]
3          [market, global, analysis, intelligence, artificial, growth, report,
players, forecast, key]
4                                [etf, stock, robotics, llc, inc, quarter, short,
daily, trust, ratings]
5                          [radio, schedule, npr, public, news, programs, donate,
air, community, music]
6  [security, cyber, market, cybersecurity, artificial, report, analysis,
global, intelligence, growth]
7            [chipsets, chip, chips, chipset, market, intelligence, edge,
artificial, report, analysis]
8                                       [news, weather, sports, local, video,
county, said, ap, ai, file]
9             [funds, distribution, fund, nav, return, capital, fiscal,
investment, distributions, ytd]

   Representative_Docs
0  [global machine learning artificial intelligence market size research report
growth trends revenue segmentation companies like aibrain amazon anki cloudminds
deepmind google etc hindaily content hindaily news space energy health world
news contact close close news space energy health world news contact categories
news global machine learning artificial intelligence market size research report
growth trends revenue segmentation companies like aibrain amazon anki cloudminds
deepmind google etc…
1  [artificial intelligence market segmentation application industry size
technology regional analysis development major companies research analysis
forecast courier thursday june breaking news wood crown moulding market global
trends growth analysis opportunities forecast artificial intelligence market
segmentation application industry size technology regional analysis development
major companies research analysis forecast airport solar panels market global
key players types applications count…
2  [healthcare artificial intelligence market key players change view global
face industry microsoft corporation alphabet inc nvidia corporation general
vision inc intel corporation ibm corporation welltok etc hindaily content
hindaily news space energy health world news contact close close news space
energy health world news contact categories politics healthcare artificial
intelligence market key players change view global face industry microsoft
corporation alphabet inc nvidia corporation ge…
3  [artificial intelligence service aiaas market predicted witness sustainable
evolution years come ksu sentinel newspaper wednesday february breaking news
global variable displacement axial piston pump market latest covid impact
analysis know brand players bosch rexroth corporation kawasaki heavy industries
fmc technologies interpump group annovi reverberi spa packaged coconut water
```

```
market increasing demand industry growth industry study pandemic future growth
analysis challenges analysis dept…
4  [short interest first trust nasdaq artificial intelligence robotics etf
nasdaqrobt declines markets daily privacy policy contact business stocks
technology health lifestyle politics daily ratings news first trust nasdaq
artificial intelligence robotics etf complete form receive latest headlines
analysts first trust nasdaq artificial intelligence robotics etf free daily
follow recent posts dcp midstream nysedcp reaches new year high santander
consumer usa nysesc hits new week high short inter…
5  [artists unhappy man submitted ai artwork contest upr utah public radio query
show noticias en español news arts culture environment health npr news politics
science spanish language news utah news arts culture environment health npr news
politics science spanish language news utah news programs upr live radio
schedule upr radio schedule upr tres radio schedule programs access utah behind
headlines eating past wild utah upr live radio schedule upr radio schedule upr
tres radio schedule progr…
6  [artificial intelligence cyber security market growth top companies trends
types application forecast news parent content top april contact disclaimer
staff privacy policy news parent bio technology market research reports bio
technology electronics emerging news finance market reports main bio technology
electronics emerging news featured intelligence cyber security market growth top
companies trends types application forecast april navanath r leave report
allinclusive research study artifi…
7  [artificial intelligence chipsets market industry players analysis new
innovation growth prospects size growth revenue development policy business
regional trends forecast technewsmobi market reports content technewsmobi market
reports reporting technology market space electric news nasa satellite climate
market forecast industry analysis market reports contact space electric news
nasa satellite climate market forecast industry analysis market reports contact
artificial intelligence chipsets…
8  [global artificial intelligence healthcare markets report growing investment
ai healthcare startups increasing crossindustry partnerships news wfmzcom
permission edit article edit close sign log dashboard logout account dashboard
profile saved items logout news coronavirus election central lehigh valley berks
regional schools world sunrise inside town espanol case missed recalls missing
persons good news weather forecast hour hour local radar news weather channel
stream river levels pocono c…
9  [virtus artificial intelligence amp technology opportunities fund discloses
sources distribution section notice florida weekenda childs homewatch livewatch
guidesouth florida uscontact usprogramming country music lifestylegray dc
tvpress releasesflood advisory effectdismiss weather alerts alerts barvirtus
artificial intelligence technology opportunities fund discloses sources
distribution section aug pm edtupdated hours agohartford conn aug prnewswire
virtus artificial intelligence technolog…
```

[49]:
```
print("Number of topics:", mod_BERT_neg.get_topic_freq().shape[0])
```

```
Number of topics: 49
```

[45]:
```
# mod_BERT_neg_reduced = mod_BERT_neg.
  ↪reduce_topics(negative_news_df['cleaned_text'].tolist(), nr_topics=150)
# print("Number of topics:", mod_BERT_neg_reduced.get_topic_freq().shape[0])
```

[50]:
```
negative_news_df['BERTopic']=mod_BERT_neg.topics_
negative_news_df['BERTopic_words'] = negative_news_df['BERTopic'].apply(lambda␣
  ↪x: mod_BERT_neg.get_topic(x))
```

[51]:
```
print(negative_news_df.shape)
negative_news_df.sample(1)
```

```
(12881, 12)
```

[51]:
```
      index        date  \
5696  85321  2021-09-26

                                                  title  \
5696  GitHub's AI programming assistant can introduce security flaws - Jimmys
Post

                                                   text  \
5696  \nGitHub's AI programming assistant can introduce security flaws - Jimmys
Post\n \n\r\n\t\tSkip to content\t\t\r\n\t\nJimmys PostTech and AI
News\n\nTech\nAdvertising\nArtificial Intelligence\nInternet
Marketing\nReviews\nBlog\n\nLatest News\n\n\n\r\n\t\t\t\t\tSeptember 26,
2021\t\t\t\t\n\nGitHub’s AI programming assistant can introduce security flaws\n
By         Matthew Sparkes\n\nSecurity flaws have been found in computer
programming code developed by AI\nmetamorworks/Getty Images\n\n\…

                                                                  cleaned_title  \
5696  githubs ai programming assistant introduce security flaws jimmys post

                                                cleaned_text  \
5696  githubs ai programming assistant introduce security flaws jimmys post
content jimmys posttech ai news tech advertising artificial intelligence
internet marketing reviews blog latest news september githubs ai programming
assistant introduce security flaws matthew sparkes security flaws found computer
programming code developed ai images neural network automatically generates
source code help human programmers complete projects found include bugs security
flaws per cent code outputs neural net…

      text_word_count  \
5696              230

                                                  lemmatized_title  \
```

```
5696  github ai program assistant introduce security flaw post

          lemmatized_text  \
5696  github ai program assistant introduce security flaw post content posttech
ai news tech advertising artificial intelligence internet marketing review blog
late news github ai program assistant introduce security flaw matthew spark
security flaw find computer programming code develop image neural network
automatically generate source code help human programmer complete project find
include bug security flaw cent code output neural basis ai programming feature
call copilot available use million…

     Sentiment  BERTopic  \
5696  Negative        -1

BERTopic_words
5696  [(ai, 0.013620617777777292), (news, 0.010600169082075632), (artificial,
0.009630297905968683), (intelligence, 0.009349346336505507), (said,
0.008330658526493099), (market, 0.007543386049473349), (technology,
0.007279096180673293), (new, 0.006806846814117346), (data,
0.006787201594589452), (use, 0.006370384251930703)]
```

[52]:
```
mod_BERT_neg.visualize_topics()
```