DLnewspaper_data

April 13, 2024

```python
[92]: """
@author: Tugba Eraslanoglu

Protest Category Prediction with Zero Shot Classification

"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```python
[72]: # Load the data and read
data_file = "DLnewspaper_data.csv"
df = pd.read_csv(data_file, encoding='utf-8', delimiter=',',␣
 ↪error_bad_lines=False)
df.head()
```

```
/tmp/ipykernel_82/3881988621.py:3: FutureWarning: The error_bad_lines argument
has been deprecated and will be removed in a future version. Use on_bad_lines in
the future.


  df = pd.read_csv(data_file, encoding='utf-8', delimiter=',',
error_bad_lines=False)
```

```
[72]:            event_date  year     month day                 mood  \
   0  2024-02-16 00:00:00  2024  February  16             Peaceful
   1  2024-02-16 00:00:00  2024  February  16             Peaceful
   2  2024-02-15 00:00:00  2024  February  15   with Intervention
   3  2024-02-15 00:00:00  2024  February  15             Peaceful
   4  2024-02-14 00:00:00  2024  February  14   with Intervention

                   district                            source  \
   0                    Mitte                             Merkur
   1                    Mitte                    Berliner Zeitung
   2                    Mitte                  Der Tagesspiegel
   3                    Mitte                 Berliner Morgenpost
   4  Friedrichshain-Kreuzberg  Berliner Morgenpost; Berliner Zeitung
```

```
       category           label  \
0   Environment   Climate Change
1    War / Peace    International
2   Environment  Transportatiton
3  Human Rights   Against Racism
4    War / Peace    International


                                             title  \
0             Ministry of Economy / Climate Activists
1  Free Navalny - Putin is a killer / Solidarity …
2                Ministry of Transport / Last Generation
3  A Sign for Democracy, Diversity, Peaceful Coex…
4          Free Palastine / Solidarity with Palastine


                                        description  crowd_size_class  \
0  About 100 climate activists participated in a …                 3
1  In the evening, about 1,000 people participate…                 6
2  3 last generation activists protested in the f…                 1
3  Actors and directors protested in berlin - mit…                 3
4  In the evening, about 230 pro-palestine demons…                 4


     crowd_size
0       100-250
1   1,000-2,500
2          0-50
3       100-250
4       250-500
```

[73]: `# Check the columns`
`df.columns`

[73]: Index(['event_date', 'year', 'month', 'day', 'mood', 'district', 'source',
       'category', 'label', 'title', 'description', 'crowd_size_class',
       'crowd_size'],
      dtype='object')

[74]: `#Check the unique sub-categories to use as candidate labels with extension`
`df['label'].unique()`

[74]: array(['Climate Change', 'International', 'Transportatiton',
       'Against Racism', 'Against War', 'Housing', 'Solidarity', 'LGBTQ+',
       'Agriculture', 'Energy', 'Labor Rights',
       'Solidarity for Prisoners', 'Education', 'Finance',
       'Animal Rights', 'Migration', 'Police Violence', 'Womens Rights',
       'Peace', 'Culture', 'Cannabis', 'Disability', 'BLM', 'Other',
       'Coronavirus', 'Nuclear', 'Drug Use', 'Traffic Accident',

```
          'Against Environment', 'Antifacist', 'Rights'], dtype=object)
```

[75]:
```python
# Main categories
df['category'].unique()
```

[75]:
```
array(['Environment', 'War / Peace', 'Human Rights', 'Local', 'Economy',
       'Health'], dtype=object)
```

[76]:
```python
# We use the description of protest to prefics the sub-categories
df['description'].head()
```

[76]:
```
0    About 100 climate activists participated in a …
1    In the evening, about 1,000 people participate…
2    3 last generation activists protested in the f…
3    Actors and directors protested in berlin – mit…
4    In the evening, about 230 pro-palestine demons…
Name: description, dtype: object
```

[77]:
```python
# We split the dataframe 1 in 5 to run the code as fast as possible. While df1␣
↪shows the first 221 line, df2 depicts the second 221 line and so on.
df1 = df.iloc[0:221]
df2 = df.iloc[221:442]
df3 = df.iloc[442:663]
df4 = df.iloc[663:884]
df5 = df.iloc[884:1107]
df1.head()
```

[77]:
```
          event_date  year     month day               mood  \
0  2024-02-16 00:00:00  2024  February  16           Peaceful
1  2024-02-16 00:00:00  2024  February  16           Peaceful
2  2024-02-15 00:00:00  2024  February  15  with Intervention
3  2024-02-15 00:00:00  2024  February  15           Peaceful
4  2024-02-14 00:00:00  2024  February  14  with Intervention

                   district                                   source  \
0                     Mitte                                   Merkur
1                     Mitte                         Berliner Zeitung
2                     Mitte                       Der Tagesspiegel
3                     Mitte                     Berliner Morgenpost
4  Friedrichshain-Kreuzberg  Berliner Morgenpost; Berliner Zeitung

        category           label  \
0   Environment    Climate Change
1   War / Peace     International
2   Environment  Transportatiton
3  Human Rights    Against Racism
4   War / Peace     International
```

```
                                          title  \
0             Ministry of Economy / Climate Activists
1    Free Navalny - Putin is a killer / Solidarity …
2                 Ministry of Transport / Last Generation
3    A Sign for Democracy, Diversity, Peaceful Coex…
4          Free Palastine / Solidarity with Palastine


                                        description  crowd_size_class  \
0    About 100 climate activists participated in a …                 3
1    In the evening, about 1,000 people participate…                 6
2    3 last generation activists protested in the f…                 1
3    Actors and directors protested in berlin – mit…                 3
4    In the evening, about 230 pro-palestine demons…                 4


     crowd_size
0       100-250
1   1,000-2,500
2          0-50
3       100-250
4       250-500
```

[78]: `df2.head()`

[78]:
```
              event_date  year   month day      mood  \
221  2023-08-26 00:00:00  2023  August  26  Peaceful
222  2023-08-26 00:00:00  2023  August  26  Peaceful
223  2023-08-25 00:00:00  2023  August  25  Peaceful
224  2023-08-24 00:00:00  2023  August  24  Peaceful
225  2023-08-24 00:00:00  2023  August  24  Peaceful


                     district  \
221  Friedrichshain-Kreuzberg
222                     Mitte
223                     Mitte
224                    Berlin
225                     Mitte


                                         source        category  \
221  Berliner Zeitung; Der Tagesspiegel; Stuttgarte…           Local
222                                    Rote Fahne   Human Rights
223                                      OhmyNews    Environment
224                                   Frankenpost    Environment
225                           Berliner Morgenpost    War / Peace


             label                                             title  \
221        Culture   zug der liebe / Mehr Liebe für die Welt von mo…
```

```
222       Migration            IGM / MLPD / issuance for miners …
223  Climate Change  korea / Fridays for Future / fukushima water r…
224  Climate Change                          Fridays for Future
225   International                     Solidarity with Ukraine


                                    description  crowd_size_class  \
221  About 10,000 people participated in the protes…                 9
222  People, including members of mlpd, igm, and si…                 3
223  People, including both ethnic korean germans a…                 3
224  Fff activists and members of other environment…                 3
225  German and ukrainian activists, including refu…                 3

         crowd_size
221  10,000-50,000
222        100-250
223        100-250
224        100-250
225        100-250
```

[79]: `df3.head()`

```
[79]:          event_date  year  month day                 mood  \
     442  2023-03-08 00:00:00  2023  March   8              Peaceful
     443  2023-03-08 00:00:00  2023  March   8  Protest with intervention
     444  2023-03-07 00:00:00  2023  March   7              Peaceful
     445  2023-03-07 00:00:00  2023  March   7              Peaceful
     446  2023-03-05 00:00:00  2023  March   5              Peaceful


                     district                              source  \
     442  Friedrichshain-Kreuzberg                    Berliner Zeitung
     443                     Mitte                    Berliner Zeitung
     444      Tempelhof-Schoneberg                  Berliner Morgenpost
     445                     Mitte                  Schwabische Zeitung
     446      Tempelhof-Schoneberg  Berliner Morgenpost; Berliner Zeitung


         category            label  \
     442  Human Rights    Womens Rights
     443  Environment    Climate Change
     444      Economy      Labor Rights
     445  Environment   Transportatiton
     446  Environment   Transportatiton


                                     title  \
     442               Internationl Womens Day
     443                       Last Generation
     444  Ver.di / BSR / higher wages amid Inflation
     445               Transport/Last Generation
```

```
446                               Transport

                                     description  crowd_size_class  \
442  Around 4,000 people protested at the u-bahnhof…                 7
443  9 alg activists blocked the danziger strasse i…                 1
444  At the call of ver.di, striking employees of t…                 3
445  4 alg activists rode a firetruck in front of t…                 1
446  Between 300 (according to the police) and 500 …                 4

        crowd_size
442  2,500-5,000
443         0-50
444      100-250
445         0-50
446      250-500
```

`df4.head()`

```
[80]:              event_date  year month day       mood          district  \
663  2022-05-23 00:00:00  2022   May  23   Peaceful              Mitte
664  2022-05-23 00:00:00  2022   May  23    Violent             Pankow
665  2022-05-23 00:00:00  2022   May  23    Violent             Pankow
666  2022-05-23 00:00:00  2022   May  23   Peaceful  Treptow-Kopenick
667  2022-05-22 00:00:00  2022   May  22   Peaceful              Mitte

                   source     category          label  \
663  Extinction Rebellion  Environment    Agriculture
664      Der Tagesspiegel       Health    Coronavirus
665      Der Tagesspiegel       Health    Coronavirus
666      Der Tagesspiegel       Health    Coronavirus
667   Berliner Morgenpost  War / Peace  International

                        title  \
663  Bayer / against pestizides
664       Grannies against Right
665          Corona Measurements
666          Corona Measurements
667       Solidarity for Ukraine

                                     description  crowd_size_class  \
663  People protested outside the bayer facilities …                 3
664  Meanwhile, around 80 people gathered at the ca…                 2
665  Around 600 people demonstrated in berlin - pan…                 5
666  Hundreds of people protested in berlin - trept…                 5
667  About 450 people protested in berlin - mitte a…                 4

     crowd_size
```

```
663       100-250
664        50-100
665      500-1000
666      500-1000
667       250-500
```

[81]: `df5.head()`

[81]:
```
            event_date  year    month day                       mood  \
884  2021-08-09 00:00:00  2021  August   9                   Peaceful
885  2021-08-08 00:00:00  2021  August   8                    Violent
886  2021-08-07 00:00:00  2021  August   7  Protest with intervention
887  2021-08-07 00:00:00  2021  August   7                   Peaceful
888  2021-08-03 00:00:00  2021  August   3                   Peaceful

                    district  \
884                    Mitte
885       Tempelhof-Schoneberg
886  Friedrichshain-Kreuzberg
887       Tempelhof-Schoneberg
888                    Mitte

                                        source      category  \
884                          Indymedia (Germany)  Human Rights
885  Berliner Morgenpost; Redaktionsnetzwerk Deutsc…         Local
886                             Berliner Zeitung   Environment
887                                Paint It Red  Human Rights
888                         Suddeutsche Zeitung  Human Rights

            label                                        title  \
884  International                    rights for indigenous people
885        Culture  kiezkultur from below - for a solidarity neigh…
886  Animal Rights                             Extinction Rebellion
887         LGBTQ+              aahainst queer transphobic attacks
888  International                          Iran / against execution

                                    description  crowd_size_class  \
884  Around 150 people protested in berlin – mitte …                 3
885  People demonstrated in berlin – tempelhof-scho…                 3
886  Around 100 extinction rebellion activists bloc…                 3
887  Around 400 people protested at luftbrucke plat…                 4
888  More than 200 people protested in berlin – mit…                 3

    crowd_size
884    100-250
885    100-250
886    100-250
```

```
887    250-500
888    100-250
```

Zero-Shot Classification

We apply to zero shot classification, which is a powerful NLP algorithm with specifically Facebook's BART Large MNLI model, to predict the protest category. We make significant strides in accurately predicting protest categories.

```
[26]: !pip install -q transformers
```

```
[29]: from transformers import pipeline

      classifier = pipeline("zero-shot-classification", model="facebook/
        ↪bart-large-mnli", device=0)
```

```
[30]: # We create a sample prediction to test the model

      sequence_to_classify = "About 100 climate activists participated in a protest␣
        ↪in front of the ministry of economy in berlin - mitte (berlin) against the␣
        ↪use of fossil fuels and the commissioning of the lng terminal in mukran. the␣
        ↪protesters demanded an immediate halt to the construction of lng␣
        ↪infrastructure in germany due to fears of possible damage to the environment␣
        ↪and tourism."
      candidate_labels = ['Climate Change', 'Agriculture','Traffic Accident', 'Animal␣
        ↪Rights', 'Against Environment',
                          'Women Rights', 'BLM', 'LGBTQ+', 'Migration','Solidarity␣
        ↪for Prisoners', 'Police Violence', 'Rights',
                          'Antifacist', 'Solidarity', 'Refugee', 'Labor Rights',␣
        ↪'Finance', 'Employee', 'Wages',
                          'Invasion', 'Conflict', 'Peace', 'Against War',␣
        ↪'International', 'Racist', 'Against Rights',
                          'Against Racism', 'Coronavirus','Disability', 'Drug Use',␣
        ↪'Cannabis', 'Medical',
                          'Housing', 'Education', 'Culture', 'Other',␣
        ↪'Transportation', 'Far_right', 'Anthropocene', 'Freedom']
      # classifier(sequence_to_classify, candidate_labels)
      result0=classifier(sequence_to_classify, candidate_labels, multi_label=True)
```

```
[31]: print(result0)
```

{'sequence': 'About 100 climate activists participated in a protest in front of
the ministry of economy in berlin - mitte (berlin) against the use of fossil
fuels and the commissioning of the lng terminal in mukran. the protesters
demanded an immediate halt to the construction of lng infrastructure in germany
due to fears of possible damage to the environment and tourism.', 'labels':
['Climate Change', 'Against Environment', 'Conflict', 'Against Rights',
'Solidarity', 'Rights', 'Freedom', 'International', 'Transportation',

'Employee', 'Far_right', 'BLM', 'Anthropocene', 'Other', 'Refugee', 'Migration',
'Coronavirus', 'Disability', 'Wages', 'Culture', 'Traffic Accident',
'Antifacist', 'Invasion', 'Peace', 'Housing', 'Solidarity for Prisoners',
'Cannabis', 'Labor Rights', 'Finance', 'Medical', 'LGBTQ+', 'Education', 'Women
Rights', 'Racist', 'Police Violence', 'Agriculture', 'Against War', 'Against
Racism', 'Drug Use', 'Animal Rights'], 'scores': [0.9590014815330505,
0.9455063939094543, 0.8928028345108032, 0.7828867435455322, 0.5106408596038818,
0.4245450794696808, 0.2945919334888458, 0.2762305736541748, 0.20066837966442108,
0.10656974464654922, 0.09933090209960938, 0.09790391474962234,
0.09288287907838821, 0.0778200775384903, 0.07172484695911407,
0.051513828337192535, 0.05124245956540108, 0.05030243098735809,
0.04099636897444725, 0.0293565783649683, 0.028862416744232178,
0.027310295030474663, 0.02619890309870243, 0.022626634687185287,
0.010906531475484371, 0.008382605388760567, 0.007868398912250996,
0.007367176469415426, 0.006443079560995102, 0.006243555806577206,
0.005656612571328878, 0.0034182416275143623, 0.003270698245614767,
0.0025355122052133083, 0.00242072844132781, 0.00220178859308362,
0.0017480816459283233, 0.0005278678727336228, 0.00029608537442982197,
0.00022978286142461002]}

According to above classification result, protest category is "Climate Change" with 0.9590 prediction score.

```
[32]:  # We start the predict the protest category for first 221 protest

       sequence_to_classify1 = df1['description']
       candidate_labels1 = ['Climate Change', 'Agriculture','Traffic Accident',
        ↪'Animal Rights', 'Against Environment',
                           'Women Rights', 'BLM', 'LGBTQ+', 'Migration','Solidarity
        ↪for Prisoners', 'Police Violence', 'Rights',
                           'Antifacist', 'Solidarity', 'Refugee', 'Labor Rights',
        ↪'Finance', 'Employee', 'Wages',
                           'Invasion', 'Conflict', 'Peace', 'Against War',
        ↪'International', 'Racist', 'Against Rights',
                           'Against Racism', 'Coronavirus','Disability', 'Drug Use',
        ↪'Cannabis', 'Medical',
                           'Housing', 'Education', 'Culture', 'Other',
        ↪'Transportation', 'Far_right', 'Anthropocene', 'Freedom']
       #classifier(sequence_to_classify, candidate_labels)
       results1=classifier(sequence_to_classify1.tolist(), candidate_labels1,
        ↪multi_label=True)
```

```
[33]:  # Create a DataFrame from results for first 221 protest

       data = []
       for i, result in enumerate(results1):
           try:
               if isinstance(result, dict):
```

```
            predicted_label = result['labels'][0] if result['labels'] else "No␣
   ↪label predicted"
            score = result['scores'][0] if result['scores'] else 0.0
            data.append([i+1, sequence_to_classify1.iloc[i], predicted_label,␣
   ↪score])
        else:
            data.append([i+1, sequence_to_classify1.iloc[i], "Result is not in␣
   ↪the expected dictionary format", 0.0])
    except Exception as e:
        data.append([i+1, "Error occurred", str(e), 0.0])

df_results1= pd.DataFrame(data, columns=['note', 'description', 'predicted␣
   ↪label', 'score'])
df_results1
```

[33]:   note   description predicted label \
     0   1 About 100 climate activists participated in a … Climate Change
     1   2 In the evening, about 1,000 people participate…   Conflict
     2   3 3 last generation activists protested in the f… Transportation
     3   4 Actors and directors protested in berlin – mit…   Peace
     4   5 In the evening, about 230 pro-palestine demons…   Conflict
     ..   …   …   …
     216   217 Members of last generation slowed down traffic… Climate Change
     217   218 At the call of the club scene, club commission…   Conflict
     218   219 At the call of the berlin anti-war coordinatio…   Against War
     219   220 Two dozen people, mostly students and school p…   Conflict
     220   221 People participated in a rally of the antimili…   Against War

             score
     0   0.959001
     1   0.881849
     2   0.986211
     3   0.692765
     4   0.736558
     ..   …
     216   0.933674
     217   0.962178
     218   0.989103
     219   0.954810
     220   0.959532

     [221 rows x 4 columns]

[34]: # We continue the predict the protest category for second 221 protest

       sequence_to_classify2 = df2['description']

```python
candidate_labels2 = ['Climate Change', 'Agriculture','Traffic Accident',
 'Animal Rights', 'Against Environment',
                     'Women Rights', 'BLM', 'LGBTQ+', 'Migration','Solidarity
 for Prisoners', 'Police Violence', 'Rights',
                     'Antifacist', 'Solidarity', 'Refugee', 'Labor Rights',
 'Finance', 'Employee', 'Wages',
                     'Invasion', 'Conflict', 'Peace', 'Against War',
 'International', 'Racist', 'Against Rights',
                     'Against Racism', 'Coronavirus','Disability', 'Drug Use',
 'Cannabis', 'Medical',
                     'Housing', 'Education', 'Culture', 'Other',
 'Transportation', 'Far_right', 'Anthropocene', 'Freedom']
#classifier(sequence_to_classify, candidate_labels)
results2=classifier(sequence_to_classify2.tolist(), candidate_labels2,
 multi_label=True)
```

```python
[35]: # Create a DataFrame from results for second 221 protest

data = []
for i, result in enumerate(results2):
    try:
        if isinstance(result, dict):
            predicted_label = result['labels'][0] if result['labels'] else "No
 label predicted"
            score = result['scores'][0] if result['scores'] else 0.0
            data.append([i+1, sequence_to_classify2.iloc[i], predicted_label,
 score])
        else:
            data.append([i+1, sequence_to_classify2.iloc[i], "Result is not in
 the expected dictionary format", 0.0])
    except Exception as e:
        data.append([i+1, "Error occurred", str(e), 0.0])

df_results2 = pd.DataFrame(data, columns=['note', 'description', 'predicted
 label', 'score'])
df_results2
```

```
[35]:      note                                    description predicted label  \
      0       1  About 10,000 people participated in the protes…         Culture
      1       2  People, including members of mlpd, igm, and si…        Conflict
      2       3  People, including both ethnic korean germans a…        Conflict
      3       4  Fff activists and members of other environment…  Climate Change
      4       5  German and ukrainian activists, including refu…         Refugee
      ..     …                                              …               …
      216   217  5 alg activists blocked the a111 highway exit …  Climate Change
      217   218  At the call of 'women, life, freedom', people …    Women Rights
```

```
218   219   Around 200, including members of kn - climate …   Climate Change
219   220   At the call of ver.di, around 6,500 (according…      International
220   221   Around 1,000 people protested in berlin to sho…        Solidarity

         score
0     0.384739
1     0.940707
2     0.957159
3     0.961713
4     0.975234
..        …
216   0.986535
217   0.985059
218   0.972534
219   0.944743
220   0.983721

[221 rows x 4 columns]
```

[36]:
```python
# We continue the predict the protest category for third 221 protest

sequence_to_classify3 = df3['description']
candidate_labels3 = ['Climate Change', 'Agriculture','Traffic Accident',
 ↪'Animal Rights', 'Against Environment',
                     'Women Rights', 'BLM', 'LGBTQ+', 'Migration','Solidarity
 ↪for Prisoners', 'Police Violence', 'Rights',
                     'Antifacist', 'Solidarity', 'Refugee', 'Labor Rights',
 ↪'Finance', 'Employee', 'Wages',
                     'Invasion', 'Conflict', 'Peace', 'Against War',
 ↪'International', 'Racist', 'Against Rights',
                     'Against Racism', 'Coronavirus','Disability', 'Drug Use',
 ↪'Cannabis', 'Medical',
                     'Housing', 'Education', 'Culture', 'Other',
 ↪'Transportation', 'Far_right', 'Anthropocene', 'Freedom']
#classifier(sequence_to_classify, candidate_labels)
results3=classifier(sequence_to_classify3.tolist(), candidate_labels3,
 ↪multi_label=True)
```

[37]:
```python
# Create a DataFrame from results for third 221 protest

data = []
for i, result in enumerate(results3):
    try:
        if isinstance(result, dict):
            predicted_label = result['labels'][0] if result['labels'] else "No
 ↪label predicted"
```

```
                score = result['scores'][0] if result['scores'] else 0.0
                data.append([i+1, sequence_to_classify3.iloc[i], predicted_label,
    ↪score])
            else:
                data.append([i+1, sequence_to_classify3.iloc[i], "Result is not in
    ↪the expected dictionary format", 0.0])
        except Exception as e:
            data.append([i+1, "Error occurred", str(e), 0.0])

df_results3 = pd.DataFrame(data, columns=['note', 'description', 'predicted
    ↪label', 'score'])
df_results3
```

```
[37]:      note                                   description predicted label  \
      0       1  Around 4,000 people protested at the u-bahnhof…    Women Rights
      1       2  9 alg activists blocked the danziger strasse i…  Climate Change
      2       3  At the call of ver.di, striking employees of t…        Conflict
      3       4  4 alg activists rode a firetruck in front of t…  Climate Change
      4       5  Between 300 (according to the police) and 500 …        Conflict
      ..     …                                               …               …
      216   217  People protested in berlin - spandau against c…     Coronavirus
      217   218  Around 800 people protested in tegel, berlin -…     Coronavirus
      218   219  People protested in berlin - neukolln against …     Coronavirus
      219   220  People protested in berlin - tempelhof-schoneb…     Coronavirus
      220   221  100 people protested in berlin - lichtenberg a…     Coronavirus

             score
      0    0.949328
      1    0.956211
      2    0.804460
      3    0.962191
      4    0.917838
      ..        …
      216  0.984753
      217  0.993488
      218  0.983264
      219  0.986797
      220  0.983471

      [221 rows x 4 columns]
```

```
[41]:  # We continue the predict the protest category for fourth 221 protest

       sequence_to_classify4 = df4['description']
       candidate_labels4 = ['Climate Change', 'Agriculture','Traffic Accident',
    ↪'Animal Rights', 'Against Environment',
```

```
                    'Women Rights', 'BLM', 'LGBTQ+', 'Migration','Solidarity␣
 ↪for Prisoners', 'Police Violence', 'Rights',
                    'Antifacist', 'Solidarity', 'Refugee', 'Labor Rights',␣
 ↪'Finance', 'Employee', 'Wages',
                    'Invasion', 'Conflict', 'Peace', 'Against War',␣
 ↪'International', 'Racist', 'Against Rights',
                    'Against Racism', 'Coronavirus','Disability', 'Drug Use',␣
 ↪'Cannabis', 'Medical',
                    'Housing', 'Education', 'Culture', 'Other',␣
 ↪'Transportation', 'Far_right', 'Anthropocene', 'Freedom']

#classifier(sequence_to_classify, candidate_labels)
results4=classifier(sequence_to_classify4.tolist(), candidate_labels4,␣
 ↪multi_label=True)
```

```
[43]: # Create a DataFrame from results for fourth 221 protest

data = []
for i, result in enumerate(results4):
    try:
        if isinstance(result, dict):
            predicted_label = result['labels'][0] if result['labels'] else "No␣
 ↪label predicted"
            score = result['scores'][0] if result['scores'] else 0.0
            data.append([i+1, sequence_to_classify4.iloc[i], predicted_label,␣
 ↪score])
        else:
            data.append([i+1, sequence_to_classify4.iloc[i], "Result is not in␣
 ↪the expected dictionary format", 0.0])
    except Exception as e:
        data.append([i+1, "Error occurred", str(e), 0.0])

df_results4 = pd.DataFrame(data, columns=['note', 'description', 'predicted␣
 ↪label', 'score'])
df_results4
```

```
[43]:     note                                       description predicted label  \
      0      1  People protested outside the bayer facilities …        Conflict
      1      2  Meanwhile, around 80 people gathered at the ca…        Conflict
      2      3  Around 600 people demonstrated in berlin – pan…     Coronavirus
      3      4  Hundreds of people protested in berlin – trept…     Coronavirus
      4      5  About 450 people protested in berlin – mitte a…      Against War
      ..   …                                                …               …
      216  217  People protested in berlin at the call of the …        Conflict
      217  218  Around 2,000 people protested in berlin – mitt…        Cannabis
      218  219  People protested at breitscheidplatz in berlin…  Against Rights
```

```
219   220   Around 40 food delivery workers and supporters…          Conflict
220   221   Activists protested in front of the brandenbur…          Conflict

          score
0     0.892165
1     0.832326
2     0.961182
3     0.981071
4     0.929869
..        …
216   0.933867
217   0.995882
218   0.943291
219   0.939149
220   0.970766

[221 rows x 4 columns]
```

[44]:
```python
# We continue the predict the protest category for fifth 221 protest

sequence_to_classify5 = df5['description']
candidate_labels5 = ['Climate Change', 'Agriculture','Traffic Accident',
 'Animal Rights', 'Against Environment',
                     'Women Rights', 'BLM', 'LGBTQ+', 'Migration','Solidarity
 for Prisoners', 'Police Violence', 'Rights',
                     'Antifacist', 'Solidarity', 'Refugee', 'Labor Rights',
 'Finance', 'Employee', 'Wages',
                     'Invasion', 'Conflict', 'Peace', 'Against War',
 'International', 'Racist', 'Against Rights',
                     'Against Racism', 'Coronavirus','Disability', 'Drug Use',
 'Cannabis', 'Medical',
                     'Housing', 'Education', 'Culture', 'Other',
 'Transportation', 'Far_right', 'Anthropocene', 'Freedom']

#classifier(sequence_to_classify, candidate_labels)
results5=classifier(sequence_to_classify5.tolist(), candidate_labels5,
 multi_label=True)
```

[45]:
```python
# Create a DataFrame from results for fifth 221 protest

data = []
for i, result in enumerate(results5):
    try:
        if isinstance(result, dict):
            predicted_label = result['labels'][0] if result['labels'] else "No
 label predicted"
```

```
            score = result['scores'][0] if result['scores'] else 0.0
            data.append([i+1, sequence_to_classify5.iloc[i], predicted_label,
  ↪score])
        else:
            data.append([i+1, sequence_to_classify5.iloc[i], "Result is not in
  ↪the expected dictionary format", 0.0])
    except Exception as e:
        data.append([i+1, "Error occurred", str(e), 0.0])

df_results5 = pd.DataFrame(data, columns=['note', 'description', 'predicted
  ↪label', 'score'])
df_results5
```

[45]:
```
      note                              description predicted label  \
0        1  Around 150 people protested in berlin - mitte …   International
1        2  People demonstrated in berlin - tempelhof-scho…         Conflict
2        3  Around 100 extinction rebellion activists bloc…  Climate Change
3        4  Around 400 people protested at luftbrucke plat…           LGBTQ+
4        5  More than 200 people protested in berlin - mit…         Conflict
..     …                                …               …
218    219  Around 610 people participated in a 'silent ma…      Coronavirus
219    220  Hundreds of people gathered for a counterprotest.         Conflict
220    221  Around 1,000 people protested in berlin, in a …         Conflict
221    222  Several hundred taxi drivers protested with a …         Conflict
222    223  More than 1,000 people gathered in berlin - ne…         Conflict

        score
0    0.977427
1    0.944802
2    0.986795
3    0.984764
4    0.891157
..        …
218  0.971213
219  0.977645
220  0.907266
221  0.973187
222  0.722058

[223 rows x 4 columns]
```

[49]:
```
# Concatenating along columns axis (axis=0)

df_predicted = pd.concat([df_results1, df_results2, df_results3, df_results4,
  ↪df_results5], axis=0)

# Save the DataFrame as CSV
```

```
df_predicted.to_csv('df_predicted.csv', index=False)
df_predicted.head()
```

[49]:   note   description predicted label \
      0   1 About 100 climate activists participated in a … Climate Change
      1   2 In the evening, about 1,000 people participate…   Conflict
      2   3 3 last generation activists protested in the f…   Transportation
      3   4 Actors and directors protested in berlin – mit…   Peace
      4   5 In the evening, about 230 pro-palestine demons…   Conflict

            score
      0 0.959001
      1 0.881849
      2 0.986211
      3 0.692765
      4 0.736558

```
# We see the predicted categories

df_predicted['predicted label'].unique()
```

[82]: array(['Climate Change', 'Conflict', 'Transportation', 'Peace',
             'Solidarity', 'LGBTQ+', 'Agriculture', 'Medical', 'Against Racism',
             'Police Violence', 'Solidarity for Prisoners', 'Education',
             'Against War', 'Employee', 'International', 'Rights', 'Refugee',
             'Animal Rights', 'Against Environment', 'Culture', 'Women Rights',
             'Cannabis', 'Antifacist', 'Coronavirus', 'Disability', 'Drug Use',
             'Wages', 'Housing', 'Traffic Accident', 'Against Rights',
             'Labor Rights', 'Other', 'BLM'], dtype=object)

```
# We see the descriptions, predicted categories(label), anbd their prediction␣
 ↪scores

df_predicted
```

[83]:   note   description predicted label \
      0   1 About 100 climate activists participated in a … Climate Change
      1   2 In the evening, about 1,000 people participate…   Conflict
      2   3 3 last generation activists protested in the f…   Transportation
      3   4 Actors and directors protested in berlin – mit…   Peace
      4   5 In the evening, about 230 pro-palestine demons…   Conflict
      ..   …   …   …
      218   219 Around 610 people participated in a 'silent ma…   Coronavirus
      219   220 Hundreds of people gathered for a counterprotest.   Conflict
      220   221 Around 1,000 people protested in berlin, in a …   Conflict
      221   222 Several hundred taxi drivers protested with a …   Conflict
```

```
222   223  More than 1,000 people gathered in berlin - ne…        Conflict

          score
0     0.959001
1     0.881849
2     0.986211
3     0.692765
4     0.736558
..        …
218   0.971213
219   0.977645
220   0.907266
221   0.973187
222   0.722058

[1107 rows x 4 columns]
```

[84]:
```python
#We apply to the metric to compare our predicted results with ground truth data
 that is created manually
from sklearn.metrics import accuracy_score, precision_score, recall_score,
 f1_score
```

[85]:
```python
# Extract predicted labels and ground truth labels

predicted_labels = df_predicted['predicted label'].astype(str)
ground_truth_labels = df['label'].astype(str)

# Calculate accuracy
accuracy = accuracy_score(ground_truth_labels, predicted_labels)

# Calculate Precision
precision = precision_score(ground_truth_labels, predicted_labels,
 average='weighted')

# Calculate Recall
recall = recall_score(ground_truth_labels, predicted_labels, average='weighted')

# Calculate F1
f1 = f1_score(ground_truth_labels, predicted_labels, average='weighted')

print(f"Accuracy is  {accuracy:.4f}")
print(f"Precision is  {precision:.4f}")
print(f"Recall is {recall:.4f}")
print(f"F1 is {f1:.4f}")
```

```
Accuracy is  0.4309
Precision is  0.6781
```

Recall is 0.4309
F1 is 0.4682

/opt/conda/lib/python3.10/site-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, msg_start, len(result))
/opt/conda/lib/python3.10/site-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Recall is ill-defined and being set to 0.0 in labels
with no true samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))

```python
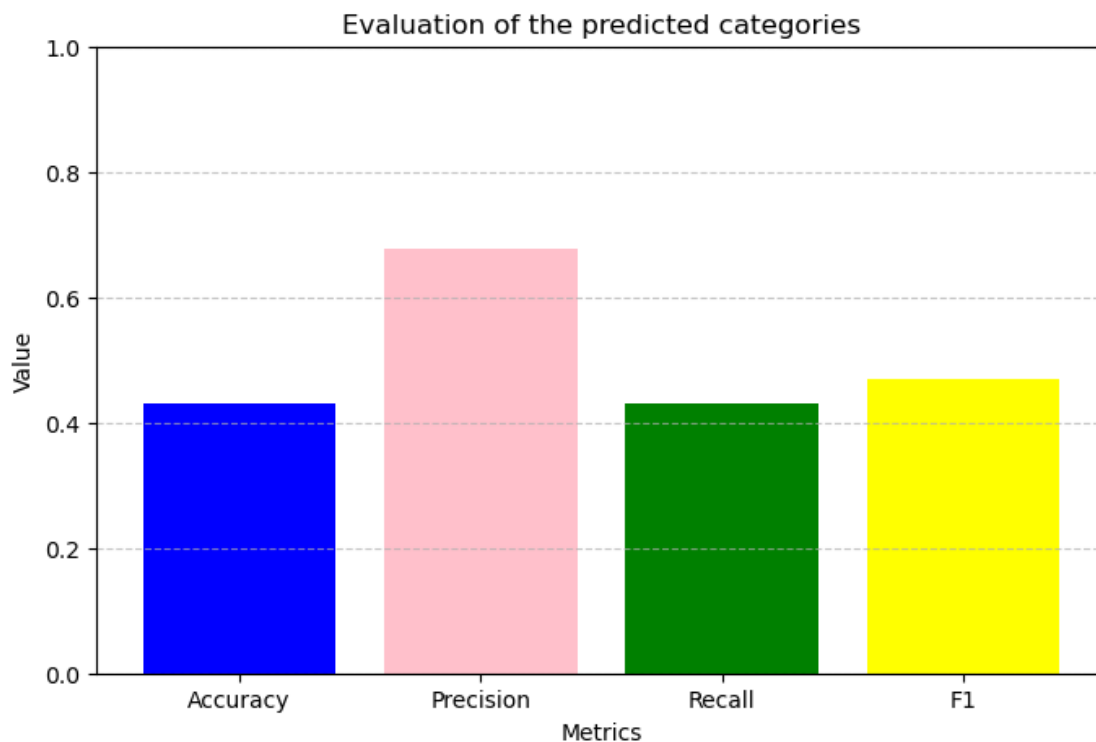[86]: # Plotting the metric values

metrics = ['Accuracy', 'Precision', 'Recall', 'F1']
values = [accuracy, precision, recall, f1]

plt.figure(figsize=(8, 5))
plt.bar(metrics, values, color=['blue', 'pink', 'green', 'yellow'])
plt.title('Evaluation of the predicted categories')
plt.xlabel('Metrics')
plt.ylabel('Value')
plt.ylim(0, 1)   # Set y-axis limit to range [0, 1]
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

```
[90]:  # Print the number of predicted categories

        predicted_label_counts = df_predicted['predicted label'].value_counts()
        print("Number of each predicted category:")
        print(predicted_label_counts)
```

```
Number of each predicted category:
Conflict                    332
Climate Change              309
Coronavirus                  90
Solidarity                   74
Against War                  33
Transportation               24
International                21
Agriculture                  20
Solidarity for Prisoners     19
Police Violence              17
Women Rights                 17
Against Racism               15
Education                    14
Housing                      11
LGBTQ+                       11
Against Environment          11
Rights                       10
Labor Rights                 10
Against Rights                9
Wages                         8
Refugee                       7
Employee                      7
Peace                         7
Culture                       6
Drug Use                      6
Cannabis                      4
Antifacist                    3
Traffic Accident              3
Disability                    2
Animal Rights                 2
Medical                       2
BLM                           2
Other                         1
Name: predicted label, dtype: int64
```

According to our findings, we can say that the number of conflict and climate change protests have
significantly increased in Berlin.

[ ]: