STEP 1. Preparing the training dataset based on Twitter

1.1.   Get the API keys

Use your Twitter account or register a new account and get API keys. It is necessary to collect information from Twitter.

After you log in, go to Settings and Privacy and create a new App using Apps pane.

1.2.   Change the source code of the file WP7_STEP1_collecting_tweets.py by adding the keys, country code and keyword you want to collect

```
# !!! put your keys in single quota in the four lines below
consumer_key = "
consumer_secret = "
access_token = "
access_token_secret = "

# !!! REPLACE pl WITH YOUR COUNTRY CODE
language='pl'

# !!! REPLACE THE keyword TO COLLECT INFORMATION BASED ON THE KEYWORD
keyword=":("
```

1.3.   Execute the file

You can execute the file by writing:

python3 WP7_STEP1_collecting_tweets.py

If you have problems with Python on your computer, just ask Jacek for help.

Repeat the step 1.3 with different keywords to create several different files for the next task – identifying tweets. The results will be stored in the files: **WP7_keyword_??????.csv**.

1.4.   Manual work – identifying tweets

Based on your collected CSV files, add additional column named **sentiment** and try to identify each tweet with the following rules, according to the European Social Survey, Social Cohesion Survey and EU Statistics on Income and Living Conditions (EU-SILC):

1) happy
2) neutral
3) calm
4) upset
5) depressed
6) discouraged
7) indeterminate

You can change the list according to your needs but for comparison purposes it would be better to collect the same list.

Based on this list the final results will be identified by Machine Learning algorithm.

STEP 2. Testing the training dataset

2.1. Combine all CSV files into one

You can just copy/paste to CSV file to have all sentiments in one CSV sheet.

No.	Tweet	Sentiment
1	Wieczorny spacer :) Coś przyjemnego na koniec dnia :D #spacer #zima #odpoczynek	happy
2	#spacer w otoczeniu pięknej śnieżnej zimy #odpoczynek	neutral
3	Trochę odpoczynku w rodzinnym mieście. #odpoczynek #weekend	neutral

Save this file as **WP7_training_data.csv** because it will be used in the next step.

2.2. Test the dataset with the training data

The goal of this step is to know the accuracy of your training dataset.

You have to execute the file like this:

python3 WP7_STEP2_testing_dataset.py

Then analyze the results, like this:

```
               precision    recall   f1-score    support

       happy       0.62       0.75       0.68         56
     neutral       0.60       0.71       0.65         34
        calm       0.43       0.30       0.35         10
       upset       0.67       0.15       0.25         13
   depressed       0.59       0.62       0.60         21
 discouraged       0.59       0.50       0.54         20
indeterminate       0.00       0.00       0.00          3

 avg / total       0.59       0.60       0.58        157
```

It means that dataset still need some enhancement for a few keywords.

STEP 3. Collecting and providing the final results

3.1. Get the current tweets from timeline

You have to execute the third script to collect the current tweets.

python3 WP7_STEP3_collecting_timeline.py

Please note that it will be stored in the file:

WP7_population_datasource.csv

3.2. Provide the results

The goal of this point is to use training dataset to classify comments from the file generated in the point 3.1.

python3 WP7_STEP3_providing_results.py