

twitter_sentiment

September 29, 2021

1 Download Twitter Data for Sentiment Analysis

We use a dataset that contains 1.6 million training and 350 test tweets from 2009 with algorithmically assigned binary positive and negative sentiment scores that are fairly evenly split.

```
[9]: from pathlib import Path
import requests
from io import BytesIO
from zipfile import ZipFile
```

1.1 Download and unzip

Download the data from [here](#).

The following code downloads and extracts the content of the compressed file and stores it in 'sentiment140', while renaming the content as follows: - training.1600000.processed.noemoticon.csv to train.csv, and - testdata.manual.2009.06.14.csv to test.csv

```
[2]: path = Path('sentiment140')
if not path.exists():
    path.mkdir()
```

```
[3]: URL = 'http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip'
```

```
[5]: response = requests.get(URL).content
```

```
[8]: with ZipFile(BytesIO(response)) as zip_file:
    for i, file in enumerate(zip_file.namelist()):
        if file.startswith('train'):
            local_file = path / 'train.csv'
        elif file.startswith('test'):
            local_file = path / 'test.csv'
        else:
            continue
        with local_file.open('wb') as output:
            for line in zip_file.open(file).readlines():
                output.write(line)
```