

```
1 ! pip install kaggle
```

↗ Show hidden output

```
1 #configure the path of kaggle.json file
2 !mkdir -p ~/.kaggle
3 !cp kaggle.json ~/.kaggle/
4 !chmod 600 ~/.kaggle/kaggle.json

1 !kaggle datasets list -s "twitter sentiment"
```

↗ Show hidden output

```
1 !kaggle datasets download -d "kazanova/sentiment140"
```

↗ Dataset URL: <https://www.kaggle.com/datasets/kazanova/sentiment140>  
License(s): other  
sentiment140.zip: Skipping, found more recently modified local copy (use --force to force download)

```
1 from zipfile import ZipFile
2 dt="/content/sentiment140.zip"
3 with ZipFile(dt,'r') as zip:
4     zip.extractall()
5     print("Extracted the file")
6
```

↗ Extracted the file

```
1 import numpy as np
2 import pandas as pd
3 import re
4 from nltk.corpus import stopwords
5 from nltk.stem.porter import PorterStemmer
6 from sklearn.feature_extraction.text import TfidfVectorizer
7 from sklearn.model_selection import train_test_split
8 from sklearn.linear_model import LogisticRegression
9 from sklearn.metrics import accuracy_score
```

```
1 import nltk
2 nltk.download('stopwords')
```

↗ [nltk\_data] Downloading package stopwords to /root/nltk\_data...  
[nltk\_data] Unzipping corpora/stopwords.zip.  
True

```
1 port_stem = PorterStemmer()
```

```
1 print(stopwords.words('english'))
```

↗ ['a', 'about', 'above', 'after', 'again', 'against', 'ain', 'all', 'am', 'an', 'and', 'any', 'are', 'aren', "aren't", 'as', 'at

```
1 twitter=pd.read_csv('/content/training.1600000.processed.noemoticon.csv',encoding='ISO.8859-1' )
```

```
1 twitter.shape
```

↗ (1599999, 6)

```
1 twitter.head()
```

				Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D
0	0	1467810672		Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
1	0	1467810917		Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
2	0	1467811184		Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
				Mon Apr 06			

```
1 column_names = ['target', 'id', 'date', 'flag', 'user', 'text']
2 twitter= pd.read_csv('/content/training.1600000.processed.noemoticon.csv',names=column_names, encoding='ISO.885
```

```
1 twitter.shape
```

```
(1600000, 6)
```

```
1 twitter.head()
2
```

	target	id	date	flag	user	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....