

Data Source and information:

You can download the training data from any of the two links:

- [Stanford link](#)
- [Google Drive link](#)

The format of the training data:

The data is a CSV with emoticons removed. Data file format has 6 fields:

- 0 - the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)
- 1 - the id of the tweet (2087)
- 2 - the date of the tweet (Sat May 16 23:58:44 UTC 2009)
- 3 - the query (lyx). If there is no query, then this value is NO_QUERY.
- 4 - the user that tweeted (robotickilldozr)
- 5 - the text of the tweet (Lyx is cool)

Source: <http://help.sentiment140.com/for-students/>

Model method references:

BOW method reference

<https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/>

Word2Vec method reference

<https://www.ahmedbesbes.com/blog/sentiment-analysis-with-keras-and-word-2-vec>