

Sentiment Analysis using NLP

Sentiment Analysis examines the problem of studying texts, like posts and reviews, uploaded by users on microblogging platforms, forums, and electronic businesses, regarding the opinions they have about a product, service, event, person, or idea.

The most common use of Sentiment Analysis is of classifying a text to a class. Depending on the dataset and the reason, Sentiment Classification can be binary (positive or negative) or multi-class (3 or more classes) problem.

The basic step for doing sentiment analysis is to collect datasets. We can either use datasets already available on the internet or can do web scraping to extract datasets from websites.

We should take care of a couple of rules before we start with web scraping:

1. You should check a site's terms and conditions before you scrape them. It's their data and they likely have some rules to govern it.
2. Be nice - A computer will send web requests much quicker than a user can. Make sure you space out your requests a bit so that you don't hammer the site's server.
3. Scrapers break - Sites change their layout all the time. If that happens, be prepared to rewrite your code.
4. Web pages are inconsistent - There's sometimes some manual clean up that has to happen even after you've gotten your data.

For web scraping we will mainly need these two libraries:

- **urlopen**

urlopen will be used to request an HTML page on the web and return its content. That's it.

- **BeautifulSoup4**

BeautifulSoup4 is a library that makes it easy to navigate in an HTML document. For example, with this library, you can easily select a table in an HTML document and iterate over its rows.

I didn't go into detail on how to do web scraping, I just made an overview for proceeding with sentiment analysis.

Once the web scraping part is done then we can proceed to pre-processing of data.

Pre-processing

An initial step in text and sentiment classification is pre-processing. A significant amount of techniques is applied to data in order to reduce the noise of text, reduce dimensionality, and assist in the improvement of classification effectiveness. The most popular techniques include:

1. Remove numbers
2. Stemming
3. Part of speech tagging
4. Remove punctuation
5. Lowercase
6. Remove stopwords

How to classify sentiment?

Machine Learning

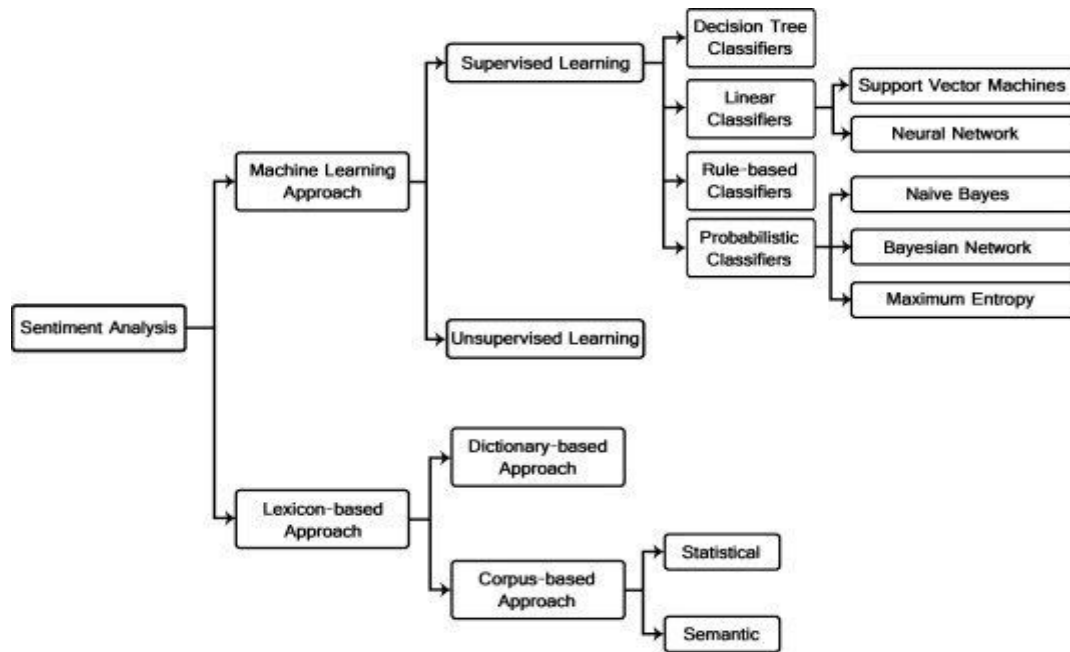
This approach, employs a machine-learning technique and diverse features to construct a classifier that can identify text that expresses sentiment. Nowadays, deep-learning methods are popular because they fit on data learning representations.

Lexicon-Based

This method uses a variety of words annotated by polarity score, to decide the general assessment score of given content. The strongest asset of this technique is that it does not require any training data, while its weakest point is that a large number of words and expressions are not included in sentiment lexicons.

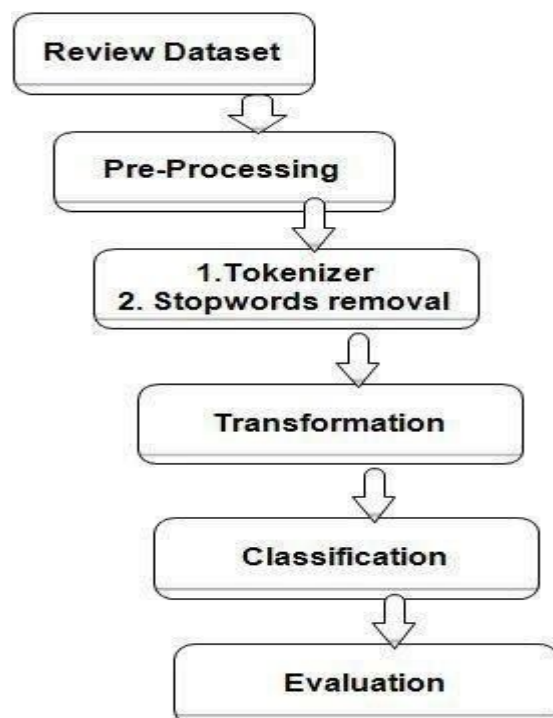
Hybrid

The combination of machine learning and lexicon-based approaches to address Sentiment Analysis is called Hybrid. Though not commonly used, this method usually produces more promising results than the approaches mentioned above.



Evaluation Metrics

As a classification problem, Sentiment Analysis uses the evaluation metrics of Precision, Recall, F-score, and Accuracy. Also, average measures like macro, micro, and weighted F1-scores are useful for multi-class problems. Depending on the balance of classes of the dataset the most appropriate metric should be used.



Visualization of results

To visualize the results of Sentiment Analysis, many people employ well-known techniques, such as graphs, histograms, and confusion matrices. Because of present multiple data domains and tasks, visualizations approaches like wordcloud, interactive maps, sparkline-style plots are also very popular.

To dive deeper into sentiment analysis I will follow these few sources:

1. [SlangSD: A Sentiment Dictionary for Slang Words](#)
2. [Mining Twitter Data with Python Part 6: Sentiment Analysis Basics](#)
3. [Political Data Science: Analyzing Trump, Clinton, and Sanders Tweets and Sentiment](#)
4. [Sentiment Analysis & Predictive Analytics for trading. Avoid this systematic mistake](#)
5. [Tutorial: Building a Twitter Sentiment Analysis Process](#)

For learning web scraping I am following:

<https://www.kdnuggets.com/2017/02/web-scraping-dataset-curation-part-1.html>

<http://www.gregreda.com/2013/03/03/web-scraping-101-with-python/>