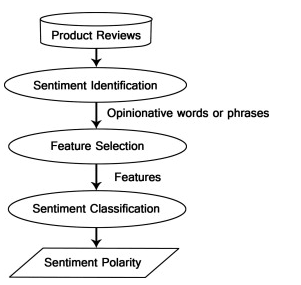
**Sentiment Analysis**

**Introduction**

Sentiment Analysis is the computational study of people’s opinions, attitudes and emotions toward an entity. The entity can represent individuals, events or topics. An opinion is an expression which consists of two key components i.e. a **target** and a **sentiment** in that particular opinion. Thus while performing sentiment analysis on an entity, one needs to identify or look for opinions in content and pick sentiments within those opinions and then classify their polarity.

For example- In “*I love this product*“, “*this product*” is the topic, and the sentiment (as expressed by the verb “*love*”) is positive.

Below is the diagram to get a clear understanding of the sentiment analysis process for product reviews



**Classification of Sentiment Analysis**

There are three main levels of classification in SA

* Document level
* Sentence level
* Aspect level

1. **Document Level**- Documents express opinions, like web reviews or open-ended questions in market surveys. And most of the times documents don’t just represent a single point of view, a single opinion. Documents typically consist of multiple opinions representing several closely related but nuanced positions. Thus the aim of document level classification is to classify an opinion document expressing a positive or negative opinion or sentiment. It considers whole document as a unit of information.

2. **Sentence level**- This classifies sentiment expressed in each sentence. Sentence-level SA will determine whether the sentence expresses positive or negative opinions. However, there is no fundamental difference between document and sentence level classifications because sentences are just short documents. Classifying text at the document level or at the sentence level does not provide the necessary detail needed opinions on all aspects of the entity which is needed in many applications, to obtain these details; we need to go to the aspect level.

3. **Aspect level**- This classifies the sentiment with respect to the specific aspects of entities. This granular analysis takes into consideration each opinion expressed in the content. The opinion holders can give different opinions for different aspects of the same entity. For example, “the voice quality of this phone is not good, but the battery life is really long”.

**Application of Sentiment Analysis**

* Product Reviews
* Stock Market
* News Article
* Political Debate (Election Candidates, Political Parties)

**Pre-Processing in Sentiment Analysis:**

**1.Noise Removal** - Cleaning the data from irrelevant information.

**2. Classification** - Categorizing the data to different domains. It is as necessary as the algorithm because you will have different set of features for different domains and thus, each domain should have different classifier.

**3. Named Entity Recognition** - This is the most important part of sentiment analysis as the objective of sentiment analysis is

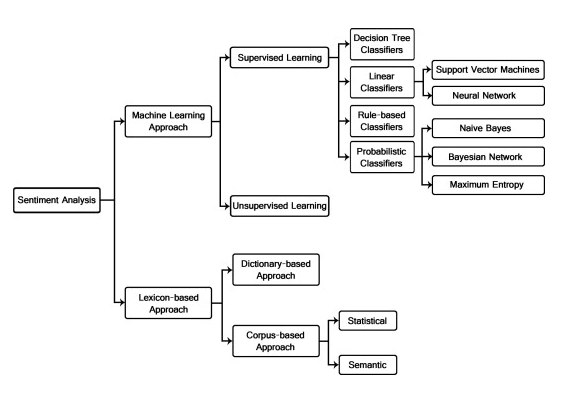
*"Given an opinion document, discover all the opinion quintuples - entity, aspect, sentiment on aspect of the entity, opinion holder and the time/context of opinion."*

**4. Subjectivity Classification** - Classifying sentences as subjective or objective since subjective sentences hold sentiments while objective sentences are facts and figures.

**5. Feature Selection** - The features can be unigrams and/or bigrams or higher ngrams with/without punctuation and with/without stop words with presence(boolean)/count(int)/tfidf(float) as accompanying feature scorer for each sentence/paragraph/file. Filtering stop words reduces accuracy. Adverbs and determiners that start with "wh" can be valuable features, and removing them as English stop words causes dip in accuracy. Similarly, punctuation helps in detecting sarcasm and exclamation.

**6. Sentiment Extraction** - It can be done using unsupervised learning, supervised learning, sentiment lexicon based approach or a mix of these.

**Classification of Sentiment Analysis**



Sentiment Classification techniques is roughly divided into machine learning approach, lexicon based approach and hybrid approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-Based Approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The Hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods.

The text classification methods using ML approach can be roughly divided into supervised and unsupervised learning methods. The supervised methods make use of a large number of labeled training documents. The unsupervised methods are used when it is difficult to find these labeled training documents. The most common algorithm used in SA are

* Naïve Bayes- computes the posterior probability of a class, based on the distribution of the words in the document. It uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label.
* Maximum Entropy- converts labeled feature sets to vectors using encoding. This encoded vector is then used to calculate weights for each feature that can then be combined to determine the most likely label for a feature set.
* Support Vector Machine- The main principle of SVMs is to determine linear separators in the search space which can best separate the different classes. Text data are ideally suited for SVM classification because of the sparse nature of text, in which few features are irrelevant, but they tend to be correlated with one another and generally organized into linearly separable categories

The lexicon-based approach depends on finding the opinion lexicon which is used to analyze the text. There are two methods in this approach. The dictionary-based approach which depends on finding opinion seed words, and then searches the dictionary of their synonyms and antonyms. The corpus-based approach begins with a seed list of opinion words, and then finds other opinion words in a large corpus to help in finding opinion words with context specific orientations. This could be done by using statistical or semantic methods.

**Issues with Sentiment Analysis**

There are four main factors that currently stops from relying blindly on tools for sentiment analysis:

1. **Context**: a positive or negative sentiment word can have the opposite connotation depending on context. For example, “my internet provider does a *great* job when it comes to stealing money from me”.
2. **Sentiment Ambiguity**: a sentence with a positive or negative word doesn’t necessarily express any sentiment. For example, “can you recommend a *good* tool I could use?” doesn’t express any sentiment, although it uses the positive sentiment word “*good*“. Likewise, sentences without sentiment words can express sentiment too. (e.g. “This browser uses a lot of memory” doesn’t contain any sentiment words, although it’s clearly negative at a document level.)
3. **Sarcasm**: a positive or negative sentiment word can switch sentiment if there is sarcasm in the sentence. For example, “Sure, I’m *happy* for my browser to crash right in the middle of my coursework”).
4. **Language**: a word can change sentiment and meaning depending on the language used. This is often seen in slang, dialects, and language variations. An example is the word “*sick*”, which can change meaning based on context, tone and language, although clear to the target audience

**References:**

1. <http://stackoverflow.com/questions/10233087/sentiment-analysis-using-r>
2. <https://lizrush.gitbooks.io/algorithms-for-webdevs-ebook/content/chapters/sentiment-analysis.html>
3. <http://saifmohammad.com/WebDocs/emotion-survey.pdf>
4. <http://analyzecore.com/2014/05/11/twitter-sentiment-analysis-based-on-affective-lexicons-in-r/>
5. <http://www.bigdatanews.com/profiles/blogs/learn-everything-about-sentiment-analysis-using-r>
6. <http://www.clarabridge.com/text-analytics-vs-sentiment-analysis/>
7. <http://andybromberg.com/sentiment-analysis/>
8. <http://www.datasciencecentral.com/profiles/blogs/text-classification-sentiment-analysis-tutorial-blog>
9. <http://datascienceplus.com/sentiment-analysis-with-machine-learning-in-r/>
10. <http://www.datasciencecentral.com/profiles/blogs/text-classification-sentiment-analysis-tutorial-blog>
11. <http://analyzecore.com/2014/04/28/twitter-sentiment-analysis/>
12. <https://www.analyticsvidhya.com/blog/2015/09/learn-top-4-hacks-perform-text-mining-faster/>
13. <https://colinpriest.com/2015/07/04/tutorial-using-r-and-twitter-to-analyse-consumer-sentiment/>
14. <http://alexloth.com/2016/01/31/how-to-implement-sentiment-analysis-in-tableau-using-r/>
15. <https://www.quora.com/What-are-some-of-the-best-ways-to-perform-a-sentiment-analysis-on-Twitter-data>
16. <http://www.dbi.io/uk/blog/sentiment-analysis-with-twitter/>
17. <https://www.youtube.com/watch?v=AZMkYst2tRM&feature=youtube_gdata><https://www.youtube.com/watch?v=dctzCcYt4AM&feature=youtube_gdata>
18. <https://blog.exploratory.io/sentiment-analysis-with-trump-clinton-sanders-twitter-data-cc978e91960f#.bhnau091x>
19. <https://www.quora.com/What-are-the-best-supervised-learning-algorithms-for-sentiment-analysis-in-text>
20. <https://www.quora.com/What-kind-of-algorithms-do-we-use-for-sentiment-analysis-Is-there-any-list-for-the-algorithms-and-about-their-structure>