**Customized rule based algorithm for sentiment analysis**

**Web Scraped Data** : Nyle shampoo reviews (242)

**Extracted from**  : Amazon

**Approach** : Divide and Conquer problem

**Abstract**:

# The customized rule based algorithm focuses on recognition of aspect term and category and classification of emotions (positive, negative, conflict, neutral) in product reviews for the aspect. So, in this I have propose the model for key phrase extraction, linguistic annotations, recognizing aspects and analyzing the sentiments for the amazon review dataset.

Keywords: Key phrase extraction, linguistic annotations, Aspects, Sentiment

Problem Statement :

* A major benefit of social media is that we can see the good and bad things people say about the particular brand or product.
* The bigger your company gets difficult it becomes to keep a handle on how everyone feels about our brand. For larger companies with thousands of daily mentions on social media, news sites and blogs, its’s extremely difficult to do this manually.
* To combat this problem, customized rule based algorithm for sentiment analysis are necessary. This rule based algorithm can be used to evaluate the people’s sentiment about particular brand or product.

Flow :

Original problem : Total Reviews are taken (242)

Sub problem 1 :

* Title and content are to be merged using separators (since both contains reviews )
* Using spacy for dependency parsing which forms the crux of aspect extraction .
* en\_core\_wb\_lg english model are used to analyse the english sentence
* replacing stop words and special characters with space
* tokenize the words and check if the competitors are mentioned in the comment (optional)
* pass the comment into spacy model
* **if token's parts of speech is noun means**
* it will c the particular noun token's immediate left (next)
* if that immediate left is compound means (next)
* suppose noun's immediate left token is compound means it (compound token and noun token) will b appended in compound pairs
* noun's immediate left -> amod (adjective modifier) and adj (adjective) present means
* noun token + immediate left amod and adj appended in amod pairs
* noun's immediate left -> advmod (adverb modifier) means
* advmod + noun are extracted and stored in amod pairs
* perform some redundancy check
* **if token's parts of speech is verb means**
* verb's immediate left is seen
* if its is advmod (adverb modifier) and adv (adverb) means
* advmod + verb is appended in advmod pairs
* if its is negative and adverb means
* negative token + adverb is extracted and stored in negative pairs
* verb's immediate right is seen
* if it is advmod (adverb modifier) and adv (adverb) means
* that string is appended in advmod pairs
* if token is adjective means
* ajectives immediate left and right is seen
* if adjective left is xcomp (open clasual complement) and adjective right is not negative means
* check adjective left -> if it is auxiliary -> auxilary verb + xcomp are extracted and stored in xcomp pairs
* or else check for adjective's immediate left xcomp and adjective's immediate negative
* negative token + auxilary +adjective are taken and stored in negative pairs
* amod\_pairs +advmod\_pairs +neg\_pairs +xcomp\_pairs are stored in pairs .
* aspect will be extracted .
* aspect key phrases are passed to lexicon to get the compound, neutral, positive, negative score .
* sentence are passed into spacy model to check the token’s are adjective or adverb or uh or jjs or rb or jj will be taken as adjective
* if token is like noun or nn or nnp or vbg or vb or vbn or nns will be taken as aspect

Sub problem 2**:**

* The reviews which is neglected will be extracted
* Removing special characters, punctuation, numbers or digits ,replacing with space
* Changing everything into lowercase
* Lemmatization is done for extracting base form of string
* Those reviews are straightly passed to lexicon to get sentiment score
* After that Key phrase are extracted
* Ranked phrased upto three are taken as top semantic phrases
* Solution of sub-problem two has been conquered.

Both solution of sub problem one and sub problem two are concatenated and solution for original problem are found .