

Sentimental Analysis with and without NLTK

Mrs.Jyoti Kolte¹, Pranjali Ghode², Swati Panday³, Khushbu Jaiswal⁴, Achal Sajanpawar⁵,

Priya Makode⁶,

1 Assistant Professor, 2, 3, 4, 5, 6Ug Students, Department of Computer Technology, Priyadarshini College of Engineering Nagpur, Maharashtra, India

I. Abstract

Sentiment Analysis is an NLP (Natural Language Processing) technique that is used to interpret the emotion behind the text. We will be able to predict if the emotion is Positive, Negative, or Neutral. Ex. Amazon: customer reviews about the product. Twitter comments help to know the majority of views.

The goal of this report is to give an introduction to this fascination project and to present a framework which will perform sentiment on sentences to analysis the NLP by associating using rule based technique and naïve bayes.

In today's era, every social media and app has widely increased data that is getting added exponentially. Every individual has their own set of information and views. To know the views about the most trending topics from a widely growing user. In this paper, we have discussed and tried to make a model based on natural language processing. Here we have discussed the two methods, that is the method using a tool kit and the method without using a tool kit. Both the methods have their own model that will work on training the set of views and words given by the user. The goal of this report is to give an introduction to this fascination project and to present a framework which will perform sentiment on sentences to analysis the NLP by associating using rule based technique and naïve bayes

III. Methodology

Rule-Based Technique: Here every word is given a value based on that the sentiments are analyzed.

Word	Sentimental Value
good	0.5
terrible	-0.8
alright	0.1

Training of a Model

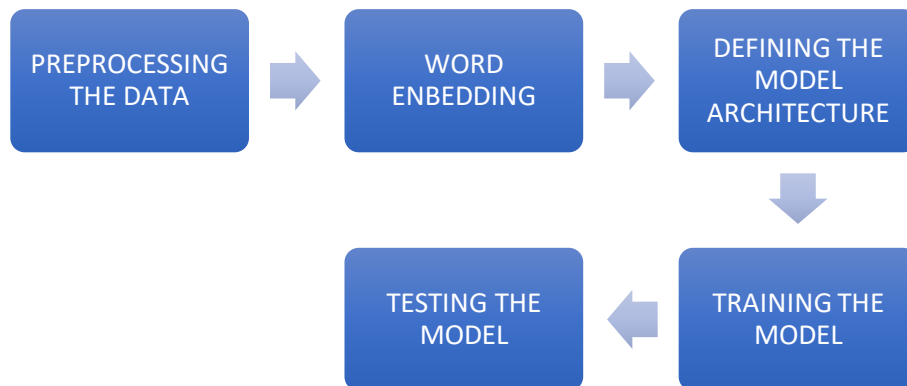


Fig: Overview of Model

CLEANING THE TEXT

- a. Creating a text file and getting a data
 - The data is then encoded into utf-8 format, as most of the data which is to analysed is in UTF(UTF-8 (UCS Transformation Format 8) is the World Wide Web's most common character encoding. Each character is represented by one to four bytes. UTF-8 is backward-compatible with ASCII and can represent any standard Unicode character.) format
- b. Converting the data into lowercase
- c. Removing the punctuations from the data
 - Converting the lowercase data into clean text without punctuation.
 - Here the ‘.translate(str.maketrans())’ removes the punctuations. It has 3 parameters. 1. First parameter specifies the list of characters that need to be replaced 2. second parameter specifies the list of chatacters with which the characters needs to be replaced, and 3. third parameter specifies the list of characters that needs to be deleted which is string.punctuation.

TOKENIZATION AND STOP WORDS

- a. Splitting text data into words
 - Splitting the sentence into words and it is stored in the form of list of words in the variable tokenized_words
 -
- b. Removing the stop words from the list of tokenized words
 - The stop words which does not add any meaning to the sentence in terms of analyzing such as I,your, as, is, the etc.. are removed from the list of tokenized words.
 - The tokenized words are compared with the data set of stop words using the iteration of for loop, if it matches with the stop words then they are removed and if not they are listed into the final_words.

EMOTION ANALYZING ALGORITHM

- Iterate through the emotions dataset, and removes all the lines and spaces using `line.replace()` and `.strip()` function
- Then the emotion meant for the word is fetched saperated by the (“ : “) semicolon and then is stored/ added into the `emotion_list` list
- The Counter is made to count the maximum number of emotions that has occurred in the `emotion_list`.

GRAPHICAL DISPLAY OF EMOTIONS

Python Package used : Matplotlib

Graph is plotted with parameters of keys and values of emotions, and kept on automatically updating mode using the function `autoformat` that will update (x,y) axis.

Which will later form an image of most dominant emotion on a graph.

TWITTER SENTIMENTAL ANALYSIS WITHOUT NLTK

Python library : GetOldTweets3

You might like the photo on the cover page as much as we do, but if it's not ideal for your report, it's easy to replace it with your own.

Just delete the placeholder picture. Then, on the Insert tab, click Picture to select one from your files.

NATURAL LANGUAGE TOOLKIT

NLTK is a standard for NPL which is used for text processing, text analytics and sentimental analysis in python.

So first we download the nltk libraries so we use its database

Tokenizing the sentence using the nltk library

importing the function `word_tokenize` from nltk library

then imporing corpus nltk stopword dataset

POSITIVE OR NEGATIVE SENTIMENT USING NLTK

The text is cleaned before tokenizing and after tokenizing and removing all the stopwords, it is then given to the function `sentimental analysis()`, and then is analysed if the emotion is positive or negative.

Here every sentiment has given a value from positive to negative.

And is shown in the graphical form.

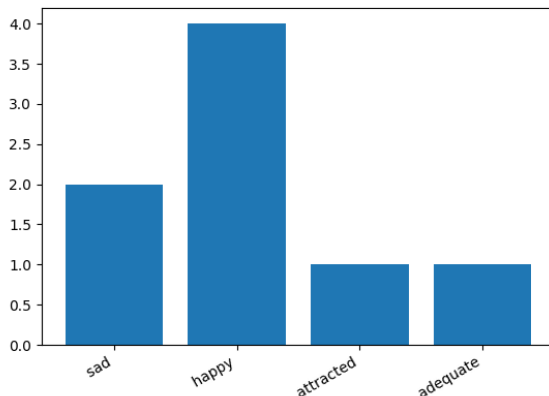


Fig: Graphical Representation

III. Result and Discussion

This is the result without natural language tool kit. Here, all the words are converted into the emotions, and counter iterates to count the number of times the emotions that has been repeated throughout the text, and performs the graphical representation as shown in the figure.

[' sad', ' happy', ' happy', ' attracted', ' happy', ' sad', ' adequate', ' happy']

Counter({' happy': 4, ' sad': 2, ' attracted': 1, ' adequate': 1})

This is the result with natural language tool kit, where it counts the most dominating emotions that is present in the text. And directly shows if the emotion is positive, neutral or negative.

Positive Sentiment .

IV. Conclusion

Above method has been applied on various review like books, tweets. We proposed a method using naïve bayes and rule based technique. We obtained an overall classification about positive and negative statement which found an accurate of over 91% on tests set over 100 reviews. The running time of our algorithm is $O(n + V \log V)$ for training and $O(n)$ for testing. Where n is the number of words in the documents (linear) and V is the size of reducing vocabulary. It is much faster than other machine learning algorithms like naïve Bayes.

V. Application of Sentimental Analysis

1. Social Media Analysis: - Social media analytics (SMA) refers to the approach of collecting data from social media sites and blogs and evaluating that data to make business decisions. This process goes beyond the usual monitoring or a basic analysis of retweets or 'likes' to develop an in-depth idea of the social consumer.

This is a pretty apt description, though we'd like to clarify that "social media sites" encompass not just Facebook, Twitter, and the like, forums and review sites, and blogs and news outlets. It's anywhere that consumers can share their beliefs, opinions, and feelings online.

2. Organization's Views: - OrganizationView is a leading People Analytics firm. We specialize in making sense of large volumes of multi-lingual text data using a human-augmented AI approach.

We provide a range of services and products to help organizations make better decisions using workforce data We view this data in its broadest sense.

3. Content Moderation: - Content moderators review user-generated content to remove offensive, inappropriate and harmful content before it reaches people online. They follow pre-defined sets of rules that platforms set to protect their users and maintain their reputation and legal compliance.

Content moderation is the process of screening and monitoring user-generated content online. To provide a safe environment for both users and brands, platforms must moderate content to ensure that it falls within pre-established guidelines of acceptable behavior that are specific to the platform and its audience.

When a platform moderates content, acceptable user-generated content (UGC) can be created and shared with other users. Inappropriate, toxic, or banned behaviors can be prevented, blocked in real-time, or removed after the fact, depending on the content moderation tools and procedures the platform has in place.

4. Stock Market Analysis: - stock market analysis is about analyzing the stock in every way possible. This obviously has a stock market analysis too. All the stocks comprise the Sensex, making it natural to study the co-dependence between the indices and each entity sectorally. In simple words, have a macroscopic viewpoint.

A Stock market analysis can be defined as an attempt to predict future market actions or to provide a general insight into the market. There are mainly two types of stock market analysis that include many concepts and factors which are to be considered when performing analysis on the stock market.

Stock Markets Analysis & Opinion Stock Markets Analysis & Opinion Morgan Stanley: Record Profits, Room to Grow Dividend Morgan Stanley (MS) is a financial holding company that offers a wide range.

5. Books review: - A book review is a form of literary criticism in which a book is merely described (summary review) or analyzed based on content, style, and merit. A book review may be a primary source, opinion piece, summary review, or scholarly review. Books can be reviewed for printed periodicals, magazines, and newspapers, as school work, or for book websites on the Internet. A book review's length may vary from a single paragraph to a substantial essay. Such a review may evaluate the book on the basis of personal taste. Reviewers may use the occasion of a book review for an extended essay that can be closely or loosely related to the subject of the book or to promulgate their own ideas on the topic of a fiction or non-fiction work.

A book review is an assessment of a particular book, usually a recent title that has just entered the market. A reviewer is providing an opinion based upon their preferences, their past reading, and overall good judgment. Ideally, a book review is honest, provides an accurate description of the book, and sums up the overall quality of the book.

VI. REFERENCES

- [1] G.Vinodhini and RM.Chandrasekaran, “**Sentiment Analysis and Opinion Mining: A Survey**”, Volume 2, Issue 6, June 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering
- [2] Liu, B.: **Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies**. San Rafael. Calif. Morgan Claypool 5(1), 1–167 (2012)
- [3] China, T.C., Joseph, S.: **A syntactic approach for aspect-based opinion mining**. In: 2015 IEEE International Conference Semantic Computing (ICSC), pp. 24–31, Feb 2015