# Topic: Text Mining (NLP)

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_Prajay B. Urkude Batch ID: 16092021**

**Topic: Text Mining and NLP**

**Problem Statement: -**

In the era of widespread internet use, it is necessary for businesses to understand what the consumers think of their products. If they can understand what the consumers like or dislike about their products, they can improve them and thereby increase their profits by keeping their customers happy. For this reason, they analyze the reviews of their products on websites such as Amazon or Snapdeal by using text mining and sentiment analysis techniques.

Task 1:

1. Extract reviews of any product from e-commerce website Amazon.
2. Perform sentiment analysis on this extracted data and build a unigram and bigram word cloud.

Ans**:-**

**Business Objective:-**

Increase the profit of the business by analyzing customers likes and dislikes about the product.

**Business Constraints:**-

Different e-commerce platform to provide the service.

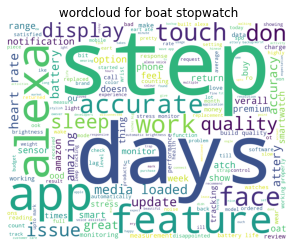
Step 1.

* Import the libraries such as request, bs4, re, matplotlib, wordcloud, nltk, sklearn
* Request: Request ia an HTTP library written in python for human being and it is used to import request to extract content from ur
* We import here BeautifulSoup function from bs4 library. BeautifulSoup is for web scraping used to scrap specific content
* We import WordCloud function from wordcloud library for creating wordcloud.

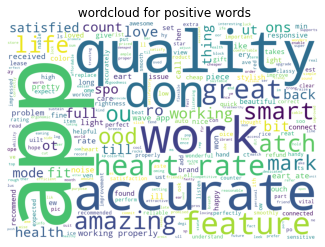
* Regular Expression re: A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

Here I am doing sentiment analysis for boat smartwatch by taking reviews from the amazon website.

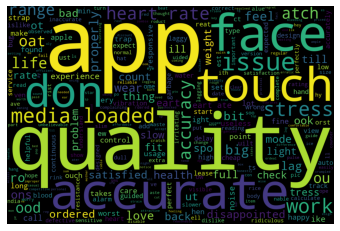
Step2

* Take the url from the website and by using the request function extract content from the url.
* Creating the soup object by extracting the specific content from the specific tag using the BeautifulSoup function and append it into the list which saves all the data which we have extracted(reviews in our case)
* Joining all the reviews in a single paragraph by using join function.
* Removing the unwanted symbols if present in the data by using sub() function of re.
* Splitting the string and them converted into sparse matrix by using Tfidf function which is extracted from sklearn.feature\_extraction.text
* Filtering the string by removing the stopwords ( which is Customed stopwords and the inbuilt stopwords.
* Draw the wordcloud by using Wordcloud function.

From the above wordcloud, step keyword is highlighted and by taking this word we see what is the review regarding this word in our string which is saved in our machine.

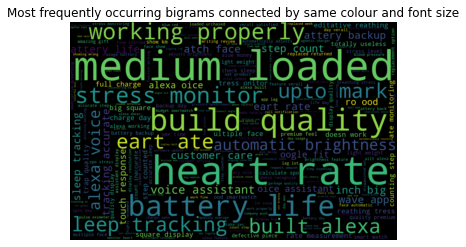
* Similarly, we create the wordcloud for positive words by taking the positive words from the strings which is in Customized list of positive words.

Here we get only positive words from which we can find out the cons of products.

* Similarly we create the wordcloud for negative words by taking the positive words from the strings which is in Customized list of positive words.

Here we get only negative words from which we can find out the cons of products.

Step3:- wordcloud for bigram

* Filtering of the string to be done by removing the special symbols, stopwords, numeric words.
* By using the bigrams function create the bigram list.
* Using countvectoriser view the frequency of bigrams.
* Generating wordcloud for bigrams

From the above wordcloud we take the heart rate as a keyword and search the review related to this keyword from the string which we have create.

Task 2:

1. Extract reviews for any movie from IMDB and perform sentiment analysis.

Ans:- **Business Objective:-**

To understand the likes and unlikes of the audience which is usful in production industry.

**Business Constraints:** -

Different e-commerce platform to provide the service.

Step 1.

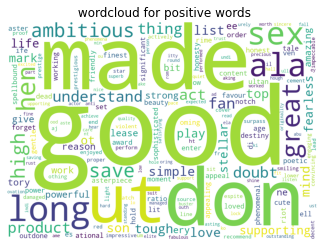
* Import the libraries such as request, bs4, re, matplotlib, wordcloud, nltk, sklearn
* Request: Request ia an HTTP library written in python for human being and it is used to import request to extract content from ur
* We import here BeautifulSoup function from bs4 library. BeautifulSoup is for web scraping used to scrap specific content
* We import WordCloud function from wordcloud library for creating wordcloud.

* Regular Expression re: A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

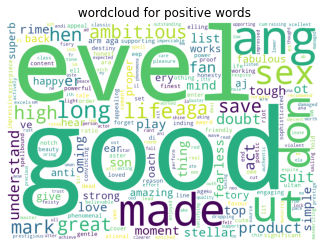
Here I am doing sentiment analysis reviews of recent movie “Gangubai Kathiyawadi” by taking reviews from the IMDB website.

Step2

* Take the url from the website and by using the request function extract content from the url.
* Creating the soup object by extracting the specific content from the specific tag using the BeautifulSoup function and append it into the list which saves all the data which we have extracted(reviews in our case)
* Joining all the reviews in a single paragraph by using join function.
* Removing the unwanted symbols if present in the data by using sub() function of re.
* Splitting the string and them converted into sparse matrix by using Tfidf function which is extracted from sklearn.feature\_extraction.text
* Filtering the string by removing the stopwords ( which is Customed stopwords and the inbuilt stopwords.

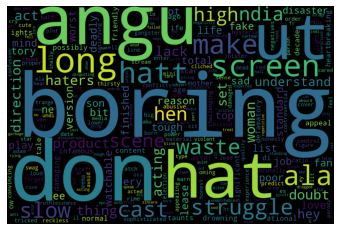
Draw the wordcloud by using Wordcloud function

From the above wordcloud, step keyword is highlighted and by taking this word we see what is the review regarding this word in our string which is saved in our machine.

* Similarly, we create the wordcloud for positive words by taking the positive words from the strings which is in Customized list of positive words.

Here we get only positive words from which we can find out the cons of products.

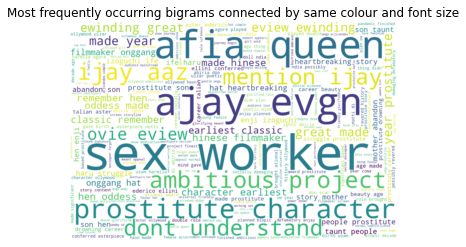
* Similarly we create the wordcloud for negative words by taking the positive words from the strings which is in Customized list of positive words.



Here we get only positive words from which we can find out the cons of products.

Step3:- wordcloud for bigram

* Filtering of the string to be done by removing the special symbols, stopwords, numeric words.
* By using the bigrams function create the bigram list.
* Using countvectoriser view the frequency of bigrams.
* Generating wordcloud for bigrams



From the above wordcloud we can make a decision how the movie is.

Task 3:

1. Choose any other website on the internet and do some research on how to extract text and perform sentiment analysis

Ans: Step 1.

* Import the libraries such as request, bs4, re, matplotlib, wordcloud, nltk, sklearn
* Request: Request ia an HTTP library written in python for human being and it is used to import request to extract content from ur
* We import here BeautifulSoup function from bs4 library. BeautifulSoup is for web scraping used to scrap specific content
* We import WordCloud function from wordcloud library for creating wordcloud.

* Regular Expression re: A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

Here I am doing sentiment analysis reviews of recent movie “Gangubai Kathiyawadi” by taking reviews from the IMDB website.

Step2

* Take the url from the website and by using the request function extract content from the url.
* Creating the soup object by extracting the specific content from the specific tag using the BeautifulSoup function and append it into the list which saves all the data which we have extracted(reviews in our case)
* Joining all the reviews in a single paragraph by using join function.
* Removing the unwanted symbols if present in the data by using sub() function of re.
* Splitting the string and them converted into sparse matrix by using Tfidf function which is extracted from sklearn.feature\_extraction.text
* Filtering the string by removing the stopwords ( which is Customed stopwords and the inbuilt stopwords.

Draw the wordcloud by using Wordcloud function

From the above wordcloud, step keyword is highlighted and by taking this word we see what is the review regarding this word in our string which is saved in our machine.

* Similarly, we create the wordcloud for positive words by taking the positive words from the strings which is in Customized list of positive words.



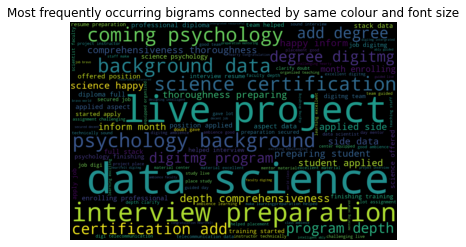
Here we get only positive words from which we can find out the cons of products.

* Similarly we create the wordcloud for negative words by taking the positive words from the strings which is in Customized list of positive words.

Here we get only positive words from which we can find out the cons of products.

Step3:- wordcloud for bigram

* Filtering of the string to be done by removing the special symbols, stopwords, numeric words.
* By using the bigrams function create the bigram list.
* Using countvectoriser view the frequency of bigrams.
* Generating wordcloud for bigrams



From the above wordcloud we can make a decision how the institution is for the data science course.