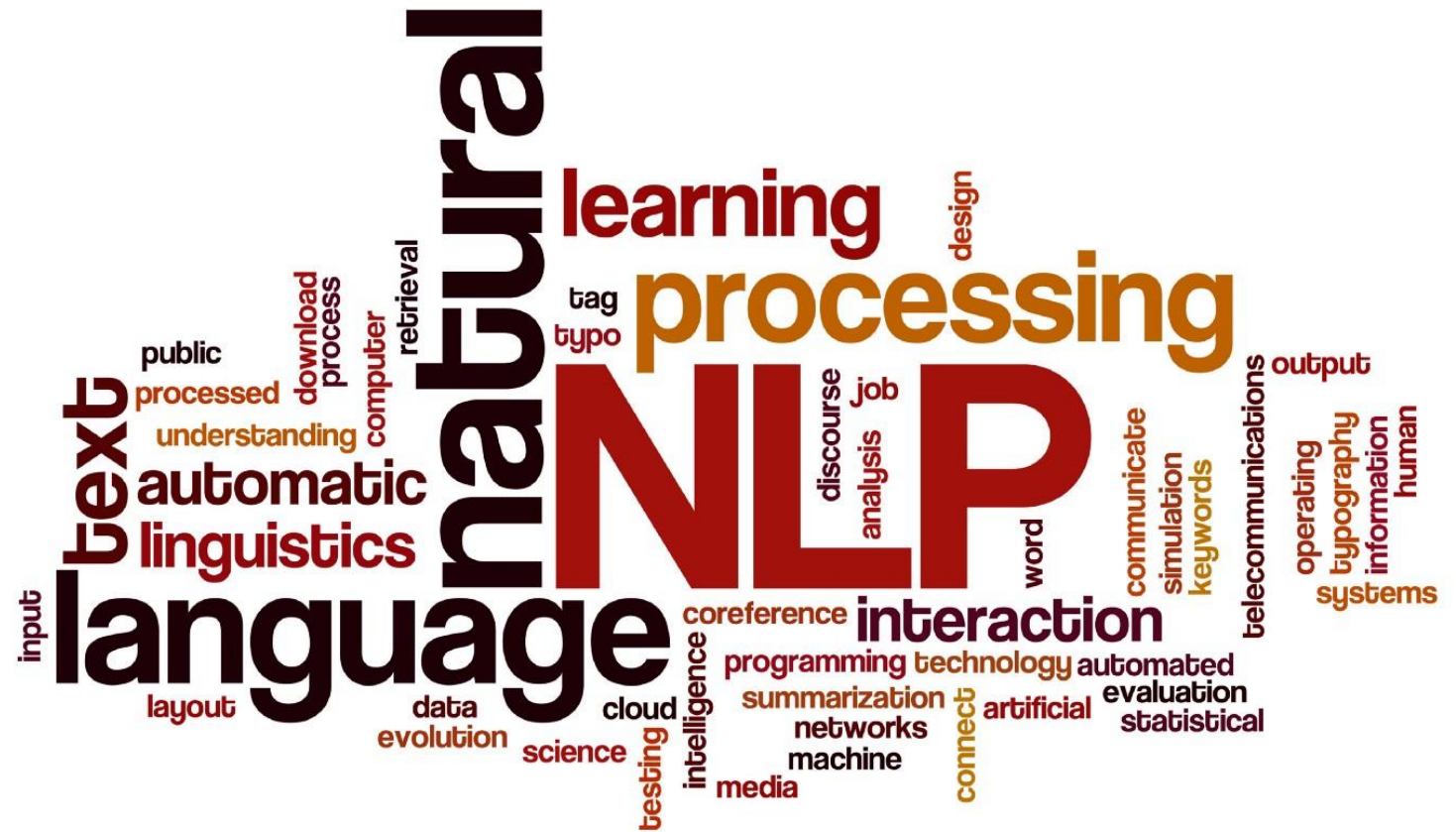


NLP Classifying positive and negative Restaurant reviews(bags of words model)



Steps involved

Import
necessary
Libraries

Import dataset

Preprocessing

Implementing
bag of words
model

Training and
Testing the
dataset

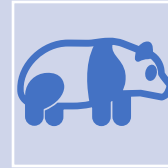
Predictive
model

Predict a single
review(positive
or negative)

Import necessary Libraries



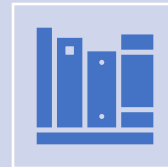
Importing necessary libraries is very important in any predicting one



Pandas is a python library which helps data manipulation and uploading dataset to our environment.



Numpy also a python library which helps in calculations .



Scikit learn is a machine learning library.

About dataset

Import dataset with setting delimiter as '\t' as columns are separated as tab space.

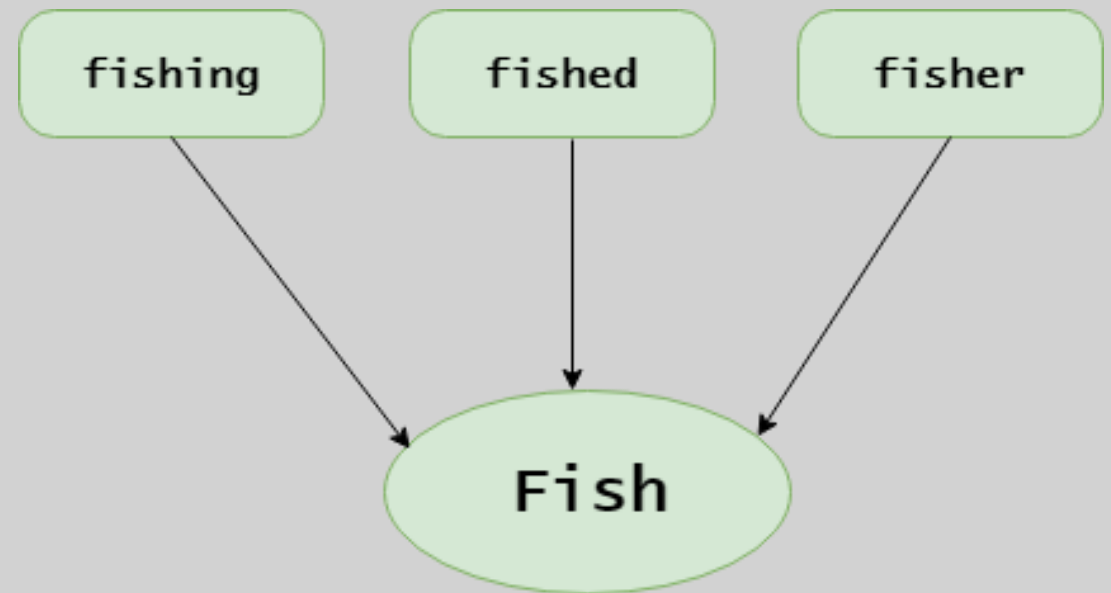
Dataset contains 1000 Reviews and two columns(Review and Liked)

Reviews and their category(0 or 1) are not separated by any other symbol but with tab space

preprocessing

- Punctuations, Numbers don't help much in processing the given text
- Stemming-Takes only root words
- Converting each word into lower case
- *Stop words are any word in a stop list which are filtered out before or after processing of natural language data*

Stemming....



Bag-Of-Words

[0, ..., 0]

20,000 elements long

171,476 words

The Second Edition of the 20-volume Oxford English Dictionary contains full entries for 171,476 words in current use, and 47,156 obsolete words. To this may be added around 9,500 derivative words included as subentries.



How many words are there in the English language?

<https://en.oxforddictionaries.com/.../how-many-words-are-there-in-the-english-language>

About this result Feedback

People also ask

How many words in the English language does the average person know? ^

Most adult native test-takers range from 20,000-35,000 words. Average native test-takers of age 8 already know 10,000 words. Average native test-takers of age 4 already know 5,000 words. Adult native test-takers learn almost 1 new word a day until middle age. May 29, 2013

Lexical facts - The Economist

<https://www.economist.com/blogs/johnson/2013/05/vocabulary-size>

Search for: How many words in the English language does the average person know?

How many words are there in all the languages in the world? v

Which language has the most words in the world? v

How many words are there in the Spanish language? v

How many words does Eminem say in a minute? v

How many pages is 100 000 words? v

- A bag of words is a sparse vector of occurrence counts of words

Implementing bag of words model

Training and Testing

To train the data. we want to divide the dataset into training and testing data.



we need CountVectorizer to set a max number of features



Do the training on the corpus(variable) and then apply the same transformation to the corpus(variable) and then convert it into an array.

Predictive model



In this predictive model we are going to implement various algorithms.



In this algorithms which one predict better accuracy consider it as a better algorithm.



In our model GaussianNB gives the accuracy of 0.73

Single review prediction

- When predicting 1000 reviews we use a loop for iteration purpose.
- In single review prediction we don't need any iteration.

