

Data preprocessing

- 1)Tokenization(Assigning values to words).
- 2) data cleaning(removing punctuation marks)
- 3)Stop words removal(Removing conjunctions or words which do not make sense).
- 4)Lemmatization(getting the root form of a word).
- 5)Stemming(Converting words into a short form, doesn't always make sense).