# Basic Terminologies in Text Mining

Bag of Words: Simplification of text. Disregarding grammar.

Corpus**:** A large set of text.

Stop Words: Common words which are not useful for deriving meaningful insights. E.g. Articles or Prepositions

Stemming**:** Different variations of a word is changed into the original root word. E.g. Chopped and Chopping is changed to Chop

Term Document Matrix (TDM): The Text/Document is present horizontally and the Term is present vertically.

Document Term Matrix (DTM): Transpose of TDM.

Term Frequency (tf): How often terms are in a document/corpus.

Lexicon**:** List of words

Bigrams**:** Collection of words taken two at a time