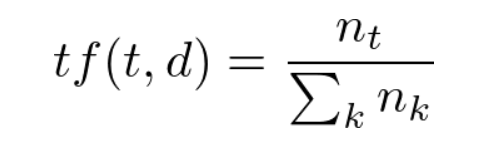
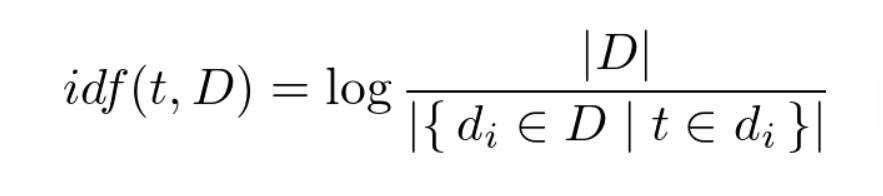
TF-IDF

* The Term Frequency-Inverse Document Frequency is a numerical statistics that demonstrates how important a word is to a corpus
* Is useful for clustering tasks, like a document clustering or it can help you understand what kind of document you have.

Term Frequency – a ratio number of current word to the number of all words in document/string/etc.



Inverse Document Frequency – a log of the ratio of the number of all documents/ string in the corpus to the number of documents with the term\_i



TF-IDF is the product tf to idf 