**Text To Numbers : Text Vectorization In NLP**

There are various methods to convert text to numbers so that machines can interpret them. The methods to convert them are as follows:

1. Binary Term Frequency -  It represents documents as binary vectors with each element indicating the presence (1) or absence (0) of a specific word in a document.

Its advantages are:

* It is simple and easy to understand text representation method.
* It is interpretable.
* It is memory efficient.

Its disadvantages are:

* It leads to information loss as it does not include frequency of any particular word even if it appears many times in a text.
* It does not capture semantic meaning of words.
* It is dependent of predefined words.It can only capture words which are already feeded in the system.

1. Bag of words – It is a widely used text vectorization technique. It represents documents as vectors where each element corresponds to the frequency of a specific term in the document.

Its advantages are:

* It is simple.
* It is versatile and can be used for various NLP Techniques.
* It is interpretable.
* It is memory efficient.

Its disadvantages are:

* No semantic understanding
* High dimensionality

1. One Hot encoding - One-Hot Encoding represents words or terms in a document as binary vectors, where each vector has a dimension for each unique term in the entire corpus.

Its advantages:

* It is simple
* It is interpretable

Its advantages:

* There is no semantic understanding
* It is memory intensive as it is highly dimensional

1. Term Frequency-Inverse Document Frequency - It combines term frequency (how often a term appears in a document) with inverse document frequency (how unique a term is across documents).

Its advantages are:

* Weighted term importance – It assigns weight to terms.
* Document Comparison - TF-IDF allows for efficient document comparison and retrieval.

Its disadvantages are:

* It is not memory efficient
* It does not provide semantic understanding.