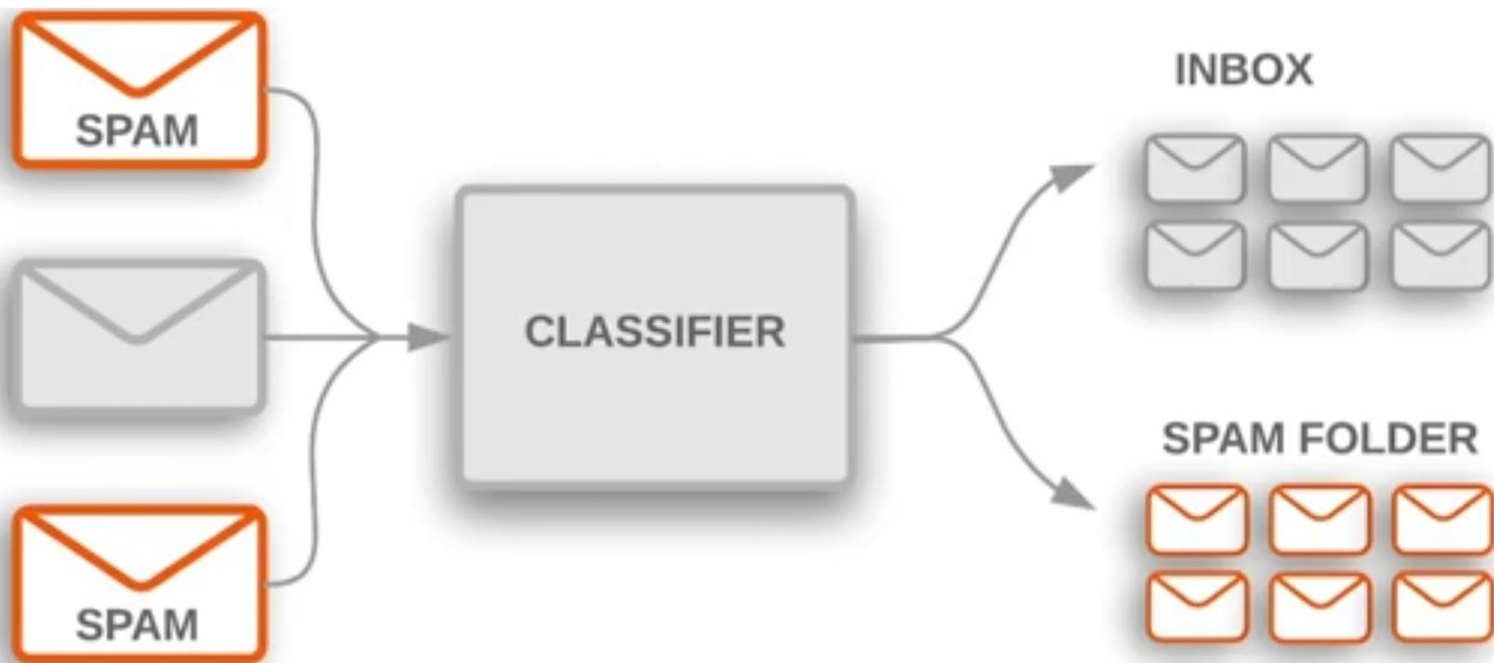


Term Frequency – Inverse Document Frequency



TF-IDF



The diagram shows the text 'TF-IDF' at the top. Two orange arrows point downwards from 'TF-IDF' to 'TF' on the left and 'IDF' on the right.

TF IDF

Term **F**requency-**I**nverse **D**ocument **F**requency

“He Loves Football”
“He is a Good Footballer”
“He is a Football Player”

Remove
Stop Words



Convert into
Lower Case

loves football
good footballer
football player

$$TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}$$

Where **t** is a **Word**
And **d** is a **Sentence**

$$TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}$$

		S1	S2	S3
loves football	loves	1/2	0	0
football lover	football	1/2	1/2	1/2
football player	lover	0	1/2	0
	player	0	0	1/2

$$\text{IDF} = \log\left(\frac{\text{No. of Sentences}}{\text{No. of Sentences Containing Words}}\right)$$

Words	Frequency		Frequency
loves	$\log(3/1)$		0.48
football	$\log(3/3)$		0
lover	$\log(3/1)$		0.48
player	$\log(3/1)$		0.48

	S1	S2	S3		Words	Frequency
loves	1/2	0	0		loves	0.48
football	1/2	1/2	1/2		football	0
lover	0	1/2	0		lover	0.48
player	0	0	1/2		player	0.48

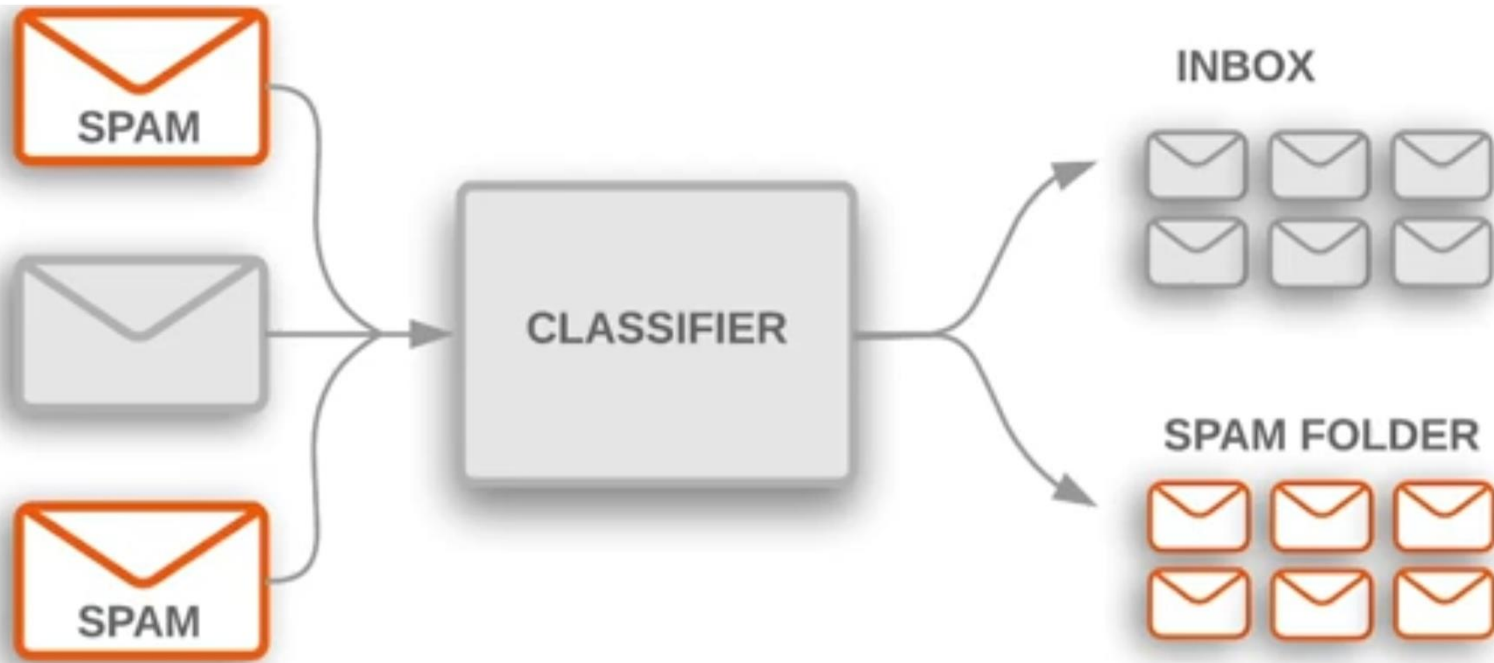
loves football
football lover
football player



loves football lover player

S1	0.24	0	0	0
S2	0	0	0.24	0
S3	0	0	0	0.24

CountVectorizer

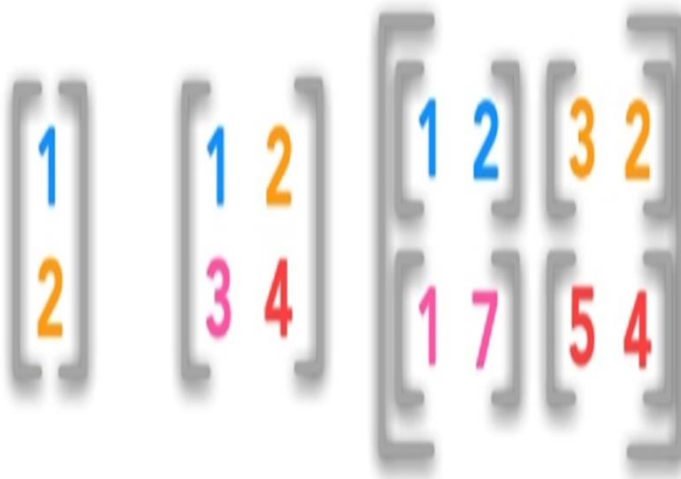


<https://towardsdatascience.com/basics-of-countvectorizer-e26677900f9c#:~:text=Countvectorizer%20is%20a%20method%20to,sparse%20matrix%20as%20shown%20below.>

https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

Scalar Vector Matrix Tensor

1



×

(11)

5 3 7

SCALAR

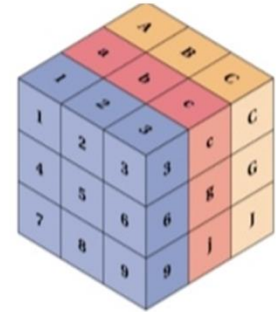
Row Vector
(shape 1x3)

5
1.5
2

Column Vector
(shape 3x1)

4 19 8
16 3 5

MATRIX



TENSOR

Understanding Scalar, Vectors, Matrices and Tensors

×

Tensors Represented by a Matrix

scalar $[a_1]$

vector $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$

dyad $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

triad

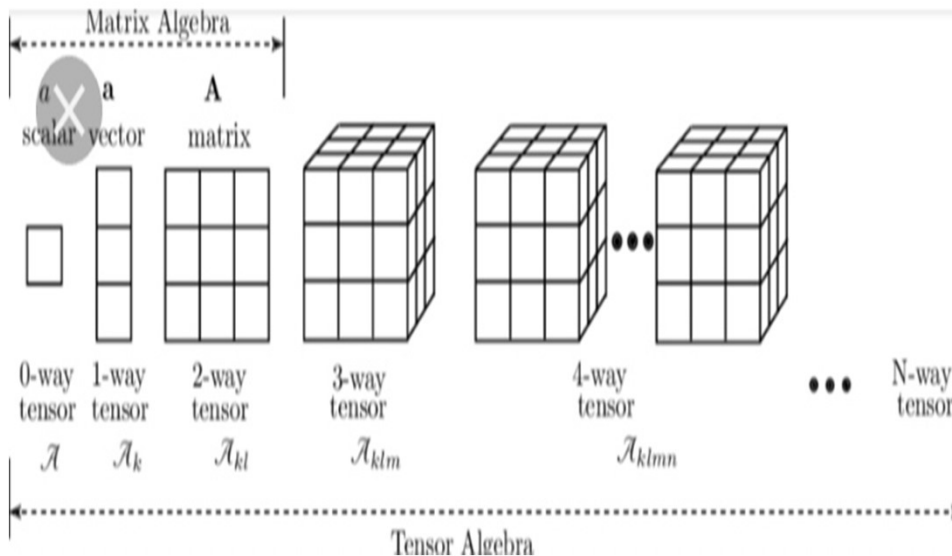
$\begin{bmatrix} a_{113} & a_{123} & a_{133} \\ a_{213} & a_{223} & a_{233} \\ a_{313} & a_{323} & a_{333} \end{bmatrix}$

$\begin{bmatrix} a_{112} & a_{122} & a_{132} \\ a_{212} & a_{222} & a_{232} \\ a_{312} & a_{322} & a_{332} \end{bmatrix}$

1st

2nd

3rd



Tensor Algebra

A vector is a **quantity or phenomenon that has two independent properties: magnitude and direction**. The term also denotes the mathematical or geometrical representation of such a quantity. Examples of vectors in nature are velocity, momentum, force, electromagnetic fields, and weight.

Speed



30Km/hr

Velocity



30Km/hr in North Direction

“Credit Card”, “Lottery”, “Make Money”

Your Credit Card Bill is



Bag of Words



"He Loves Football"
"He is a Good Footballer"
"He is a Football Player"

Remove
Stop Words



Convert into
Lower Case

loves football
good footballer
football player

Words Frequency

loves football
good footballer
football player



loves	1
football	2
good	1
footballer	1
player	1

Words Frequency

loves	1
football	2
good	1
footballer	1
player	1



	Loves	Football	Good	Foot Baller	Player
S1	1	1	0	0	0
S2	0	0	1	1	0
S3	0	1	0	0	1