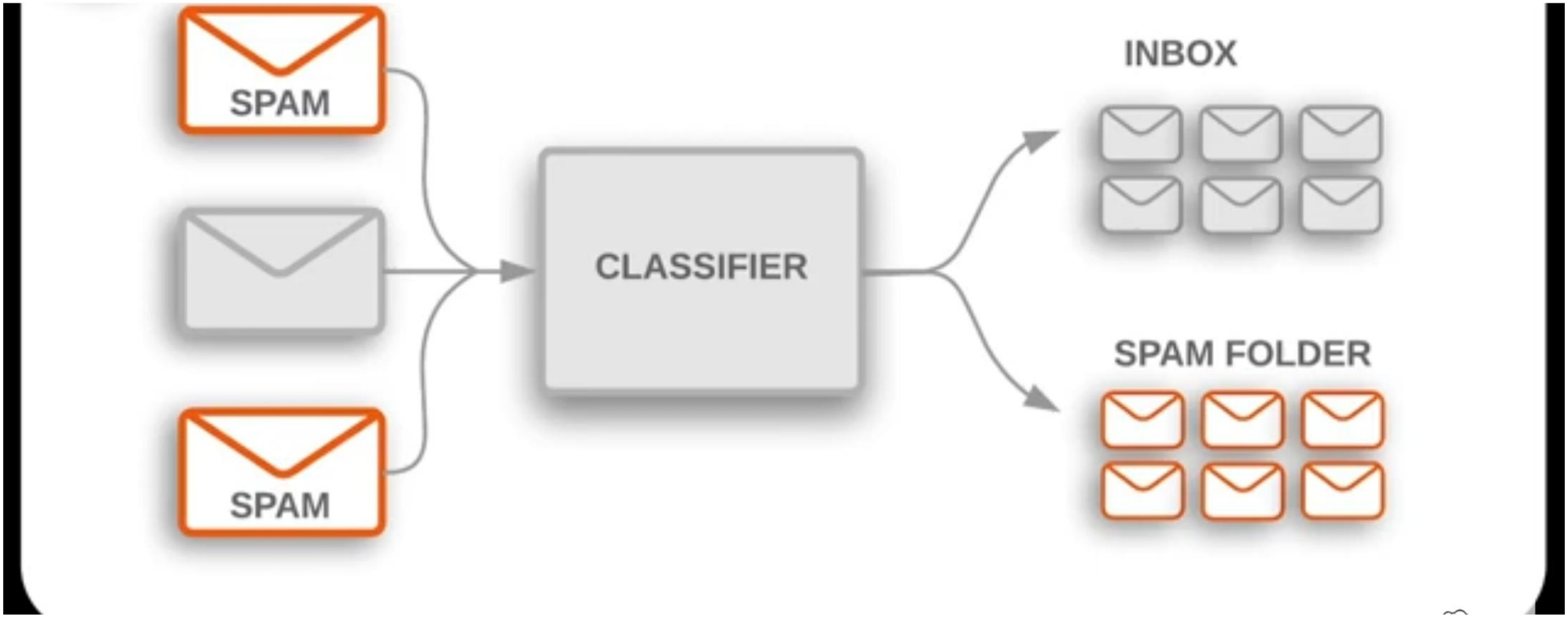
Term Frequency – Inverse Document Frequency



TF-IDF

TF IDF

Term Frequency-Inverse Document Frequency

"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

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

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

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

loves football
football lover
football player

	S1	S2	S3
loves	½	0	0
football	½	½	½
lover	0	½	0
player	0	0	½

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

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
loves	½	0	0
football	½	½	½
lover	0	½	0
player	0	0	½

Words	Frequency
loves	0.48
football	0
lover	0.48
player	0.48

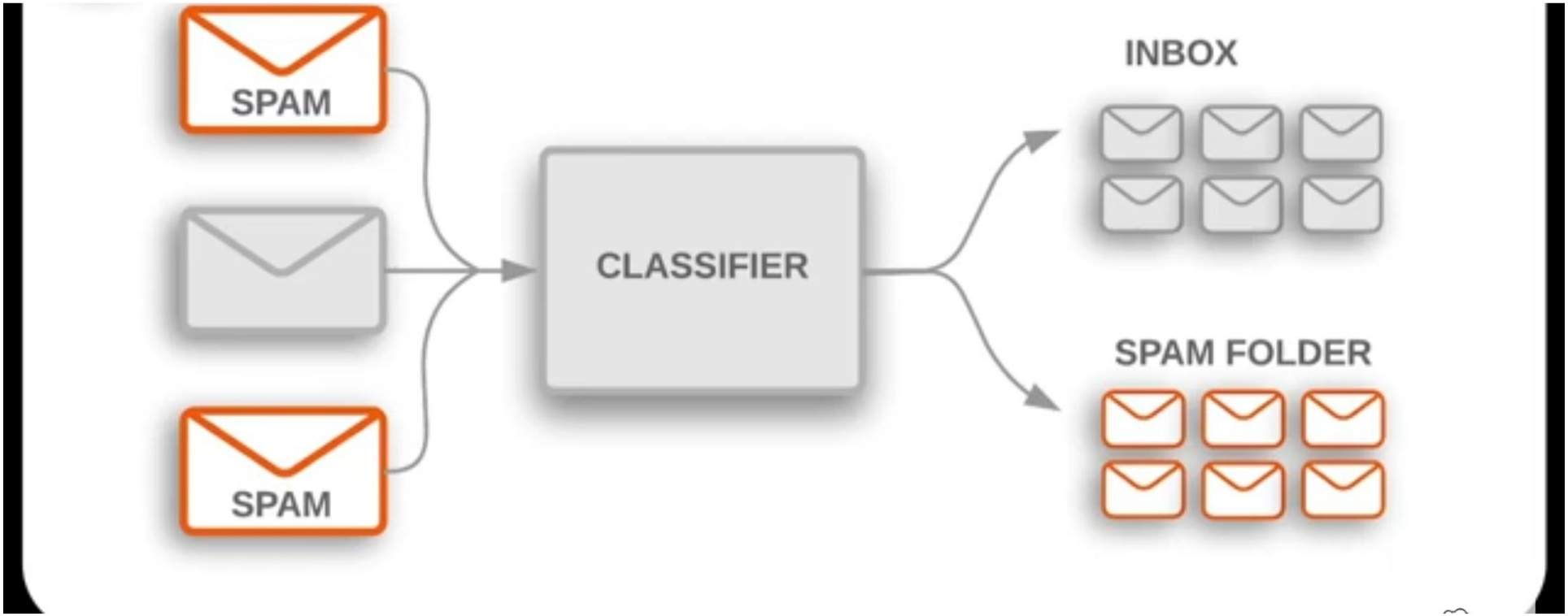
loves football

football lover

football player

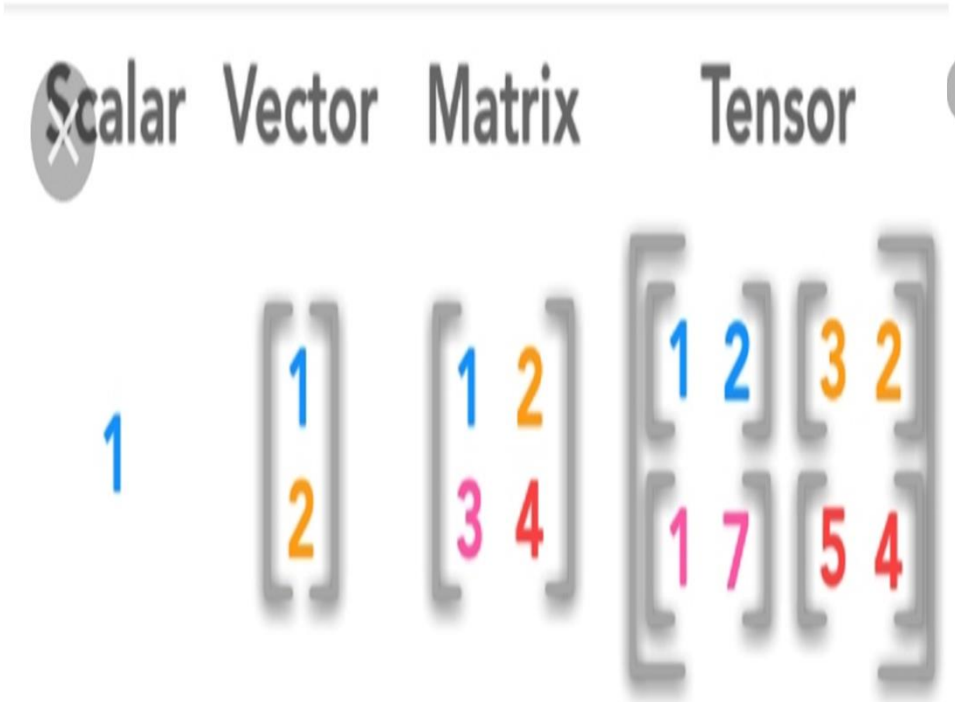
	loves	football	lover	player
S1	0.24	0	0	0
S2	0	0	0.24	0
S3	0	0	0	0.24

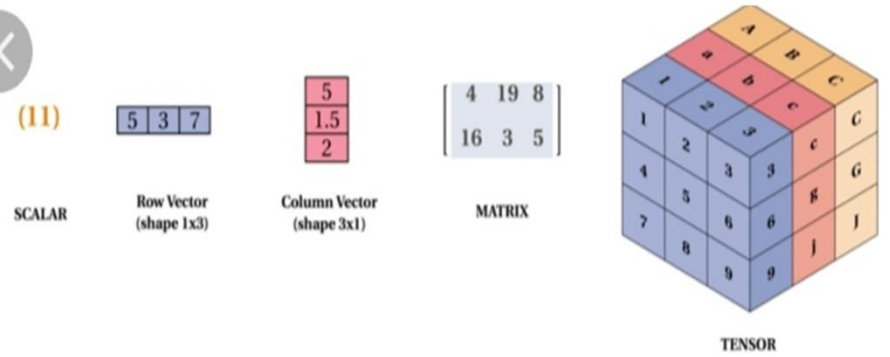
CountVectorizer



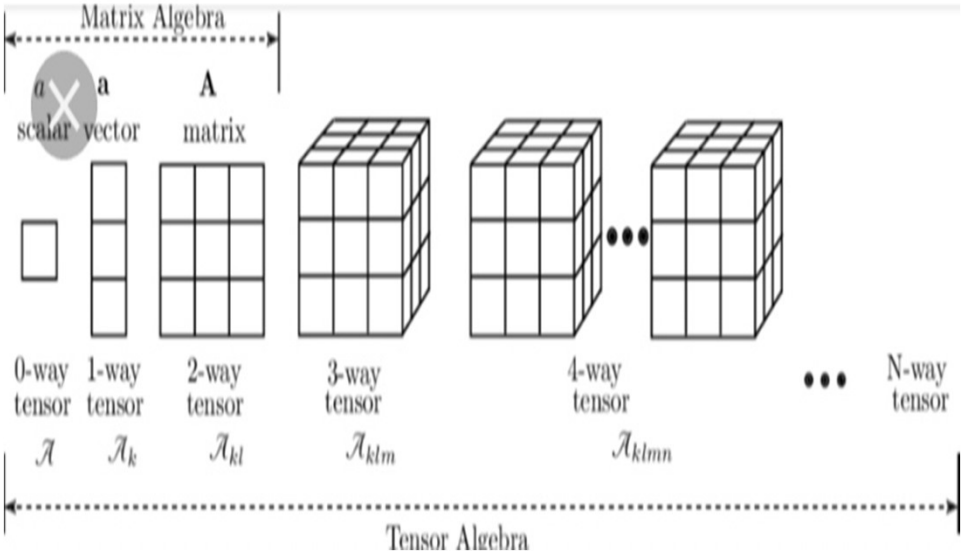
https://towardsdatascience.com/basics-of-countvectorizer-e26677900f9c#:~:text=Countvectorizer%20is%20a%20method%20to,sparse%20matrix%20as%20shown%20below.

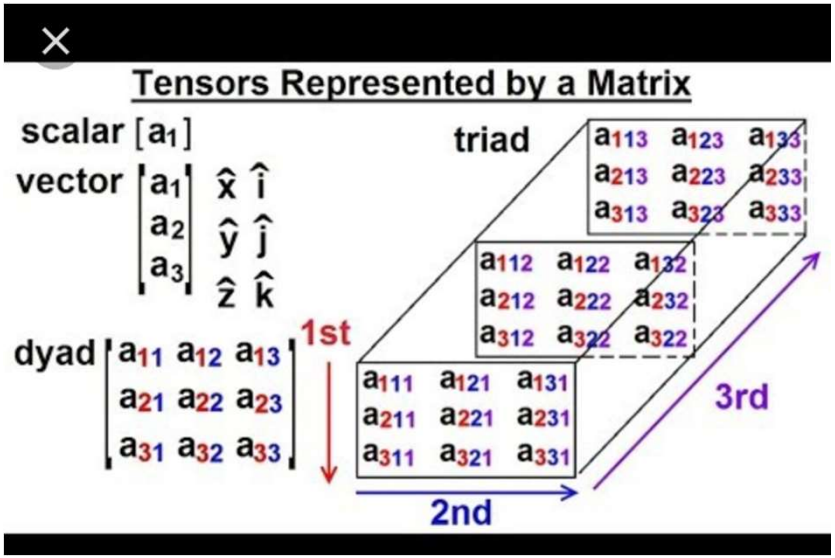
https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html





Understanding Scalar, Vectors, Matrices and Tensors





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

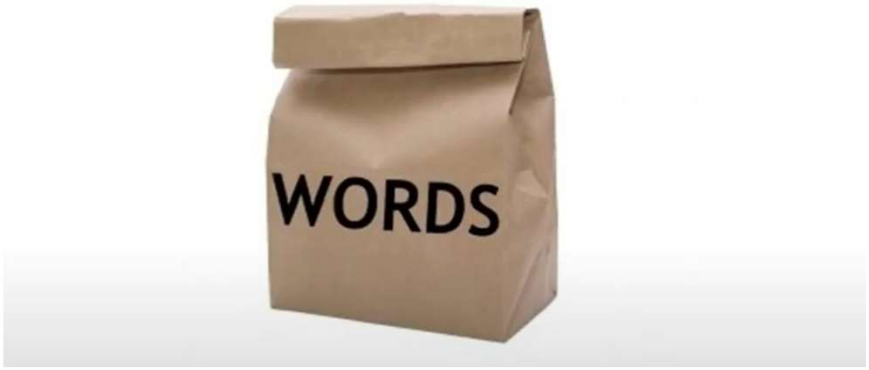
"Credit Card", "Lottery", "Make Money"

Your Credit Card Bill is .....

amazon

HDFC BANK

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

loves football
good footballer
football player

Words	Frequency
loves	1
football	2
good	1
footballer	1
player	1

Words	Frequency
loves	1
football	2
good	1
footballe	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