② **Bag of Words :- (BOW)**

Dataset

Text	O/P
He is a good boy	1
She is a good girl	1
Boy and girl are good	1

⤵ **Text preprocessing**
(Lower all the words case
Stop word)

Vocabulary

good boy girl

frequency
  3   2   2 .

$S_1 \rightarrow$ good boy
$S_2 \rightarrow$ good girl
$S_3 \rightarrow$ Boy girl good

	good	boy	girl	O/P
S_1	1	1	0	1
S_2	1	0	1	1
S_3	1	1	1	1

Binary bag of Words : $\{$ either 1 or 0 $\}$ & **BOW**

$\{$ count will get updated based on frequency $\}$.

○ **Advantages :**

→ Simple and Intuitive
→ fixed sized I/P → ML Algorithm.
  Apply कर सकते हैं!!

○ **Dis Advantages**

① Sparse Matrix → Overfitting.
② Ordering of the word is getting changed.
③ Out of Vocabulary. (OOV) .

④ Semantic meaning is not getting captured.

ent → the food is Good → [1 1 1 0 1] }
→ The food is not Good → [1 1 1 1 1] } ⇒ Sentences are almost similar.
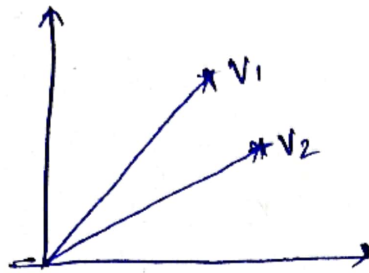
example :

S1 → I love cricket

S2 → I love football.

Vocabulary	I	love	cricket	football
	1		1	1
Frequency	2	2	1	1

	I	love	cricket	football	o/p
S1 →	1	1	1	0	1
S2	1	1	0	1	1

Simple Means :- "Bag of words ek model hai jo Text Ko words ki frequency ke basis per numbers me convert karta hai Taki ML Algorithm use samjh ske !! #