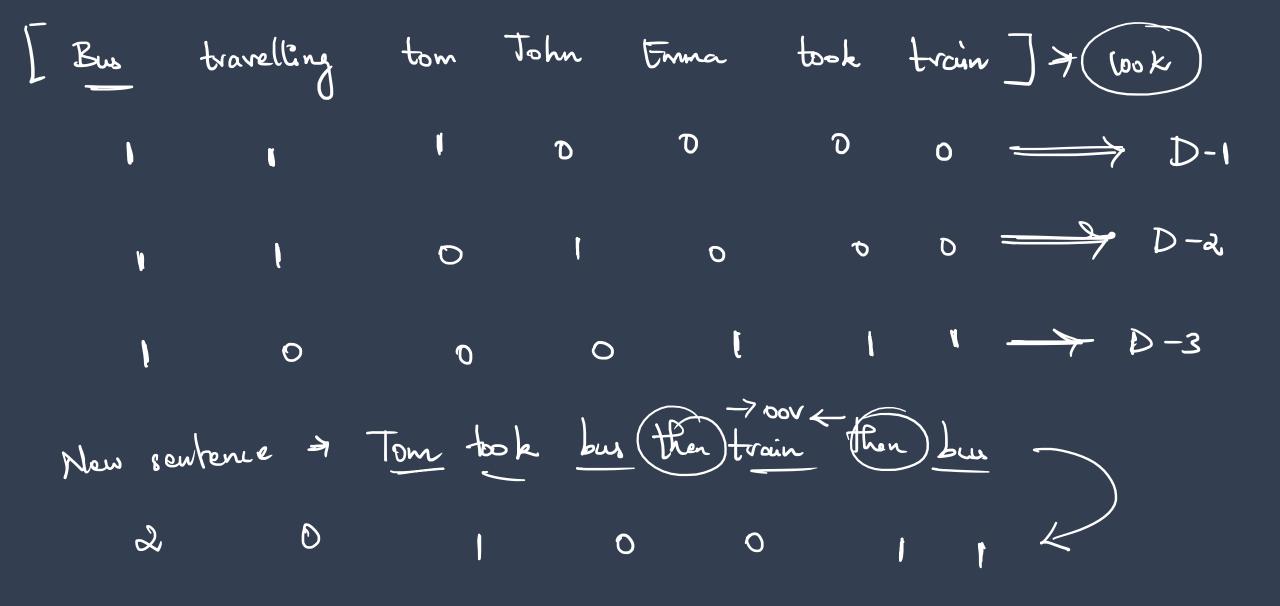
Bag of Words: (BOW) > frequency of words DI 7 Tom travelling bus. Tom "is travelling in bus D2 > John travelly bu.

D3 > 5mme took train bu John is travelling in bus Emma took train and bus [Intoods are arranged in descending order of fraguency and kept Vocabulary: John - 1 Tom - 1 in the bag. (ix) Emma - 1 travelliny - 2 Lbu, travelling, tom, john, took - 1 Lus - 3 Emna, took, frain] train - 1



## D'isadvantage:

ok sparse matrix

de bloods ave shuffted baredon frequency.

a semantic meaning is unclear.

L DOV.

#### Advantage.

& No need of fixed bought document.

de Easy implementation & out effective.

TF-IDF:

[Frequency - Inverse Document Frequency]

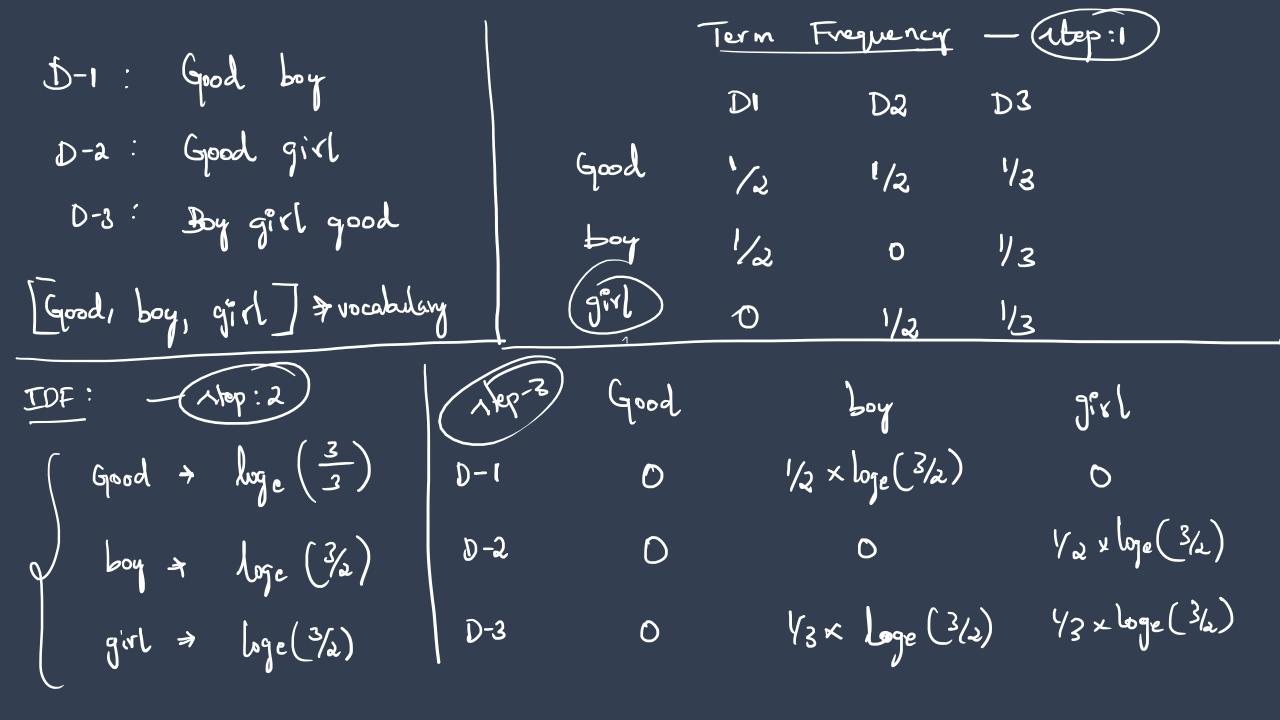
Term Frequency: Focused on Words in each sentence

# repotation of words in sentence

# words in the sentence

(F) Inverse document Frequency: Focused on the rentences in the corpus.

IDF: loge # rentences containing the word



### Disadrantage

7 004

7 sparse matrix

#### Advantage:

y word importance is captured.

7 length of rentences can be dynamic.

7 Fang to implement.

# Word Frequency and encoding: 7 one hot encoding 7 Dag of Words (Bow) Common Issue æ sparse matrix → TF-IDF 4 00V To fix the above issue Google come up with the ideal (x.) of deep learning trained library in 2013 SANN Whordavec - CROW (continious bag of words)

bing Ducen Apple Women ⇒ Boy Girl Mango Vocabulary D.02 0.98 0.01 Gender -0.97-0.01 99.0 0.02 Royal 0.99 0.01 0.03 0.4 045 Age 0.9 0.91 0.9 0.89 Food 0.96 0.01 0.01 0.02 0.02 **J** → Gender [king - boy + Rnoven] = Woman amentions A Gender [0.28 - 1 - 0.97] → 0.99 (oov is fixed)

