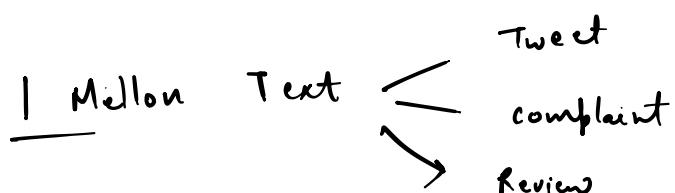
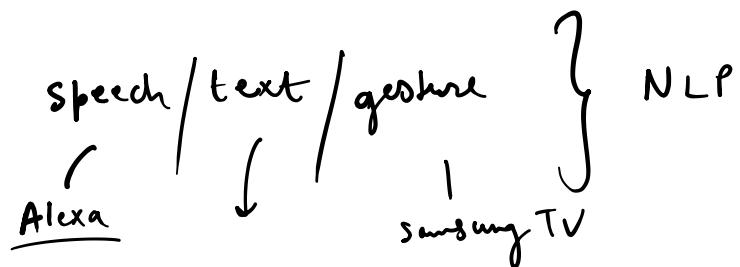


Assignment   No of questions:														
ALL QUESTIONS		S1 > Q9	Q1	Q3	Q2	Q8	Q5	Q6	Q7	Q4	Q11	Q10	0 / 11 Attempted	
<b>S1: Default Section</b>														
#	QUESTION	DIFFICULTY	TYPE								BATCH PSP	OVERALL PSP		
9	drawback of bow	Easy	☒	Objective							Not Attempted	Solve	21.9	21.65
1	ML Paradigm	Easy	☒	Objective							Not Attempted	Solve	38.1	36.36
3	Applications of Unsupervised learning	Easy	☒	Objective							Not Attempted	Solve	38.1	35.5
2	Applications of the Supervised Learning	Medium	☒	Objective							Not Attempted	Solve	36.19	34.63
8	BOW vector	Easy	☒	Objective							Not Attempted	Solve	23.81	23.81
5	Poseidon	Easy	☒	Objective							Not Attempted	Solve	36.19	35.06
6	Price is right	Easy	☒	Objective							Not Attempted	Solve	35.24	32.9
7	Person of Interest	Easy	☒	Objective							Not Attempted	Solve	34.29	32.9

/ \  
 Agenda  
 Bow  
 TF IDF — wordvec.  
 Text processing  
 Lower  
 stemming  
 Lemmatization  
 stop  
 wordcloud  
 sentiment analysis.

N L P  
 |  
 Natural Language Processing

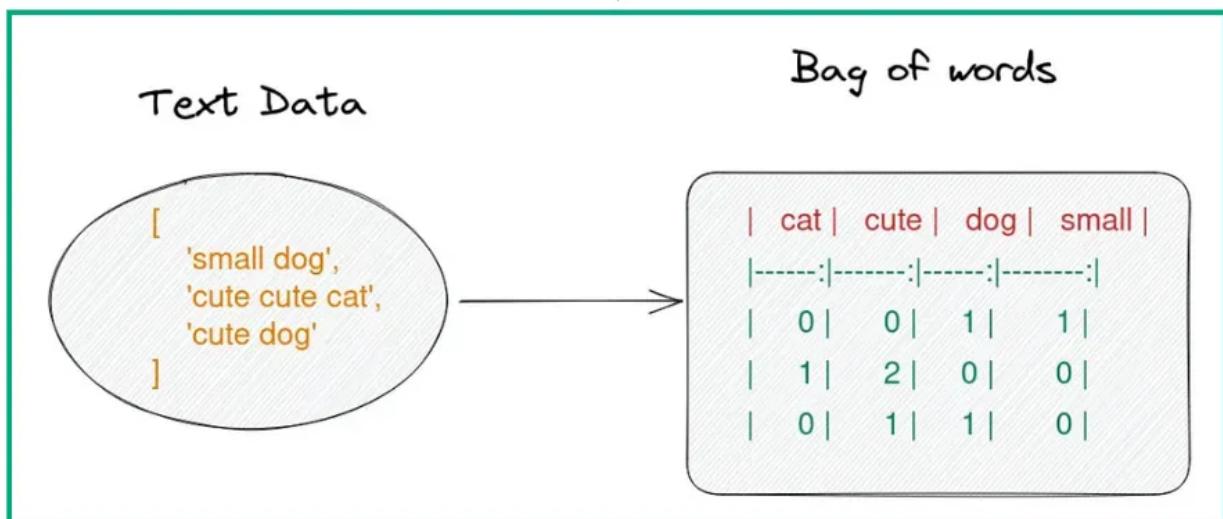
Speech  
 text  
 gesture. }  
 Natural Lang



① Text to understand

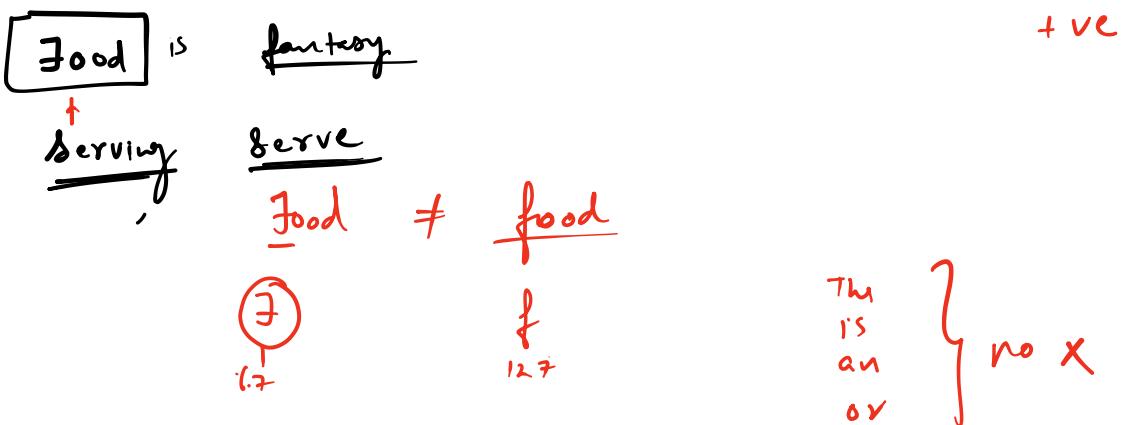
Q     " This class is     "   
                awesome.

words → ~~X~~  
AKASH  
↓↓



The food is good in this restau—  
+ve  
-ve

The service is worst



- ① .lower()
- ② .stopwords (remove)
- ③ playing → play Root word

Grammacy

Sentence → Numbers

~~The~~ food ~~is~~ good ~~in~~ ~~this~~ ~~resta-~~ food  
~~The~~ service ~~is~~ worst

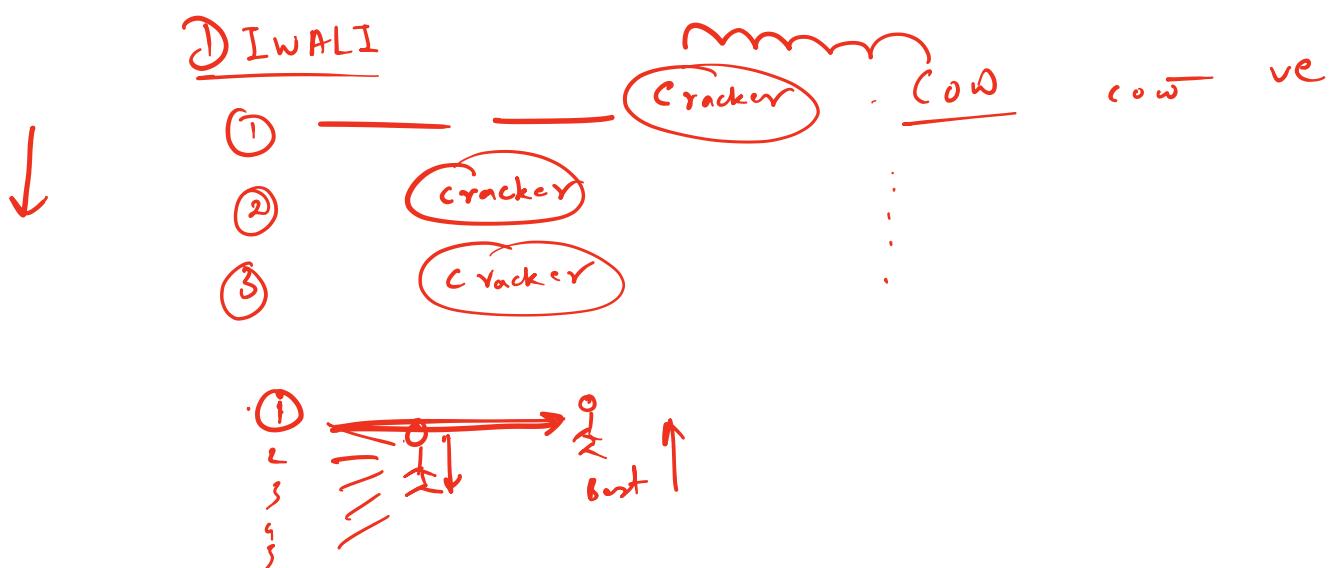
Food ~~is~~ fantoy

- ① Vocabulary (Token)
- ② Count

	food	good	service	worst	fantoy.	resta-
①	2	1	0	0	0	1
②	0	0	1	1	0	0
③	1	0	0	0	1	0

- ⑥ Drawback  
order does not matter.

① Every word has equal imp.



① rock (Token)

food      Decli . - . - . - .

TF Term frequency  $\rightarrow$  count the term  
Total words

$$TF(\text{food}) = \frac{1}{4}$$

$20 \rightarrow A1 \quad A1 \quad A1$

6

If word appears in a doc multiple times, the doc is about that word. 

① AI  
② AI

10

Pros

negate / un

$$\frac{N - \text{Total Doc}}{\text{# of doc word is after}}$$

$$\frac{10}{1}$$

$$\frac{10}{1}$$

$$\frac{1}{4} \times \frac{10}{1}$$

$\frac{1}{4} \times \frac{10}{1}$   
Invert doc freq.

$$\frac{1}{4} \times 10$$

$$\frac{1}{4} \times \frac{10}{10}$$

Deno ↑

$$1/n \times 1$$

Reduced

$$\frac{1}{\text{Doc 1}} \leftarrow \frac{2/7 \times 3}{\text{TF} \times \text{IDF}}$$

2  
Doc 2  
Doc 3

$w_1 \quad w_2 \quad w_3 \quad w_4 \quad w_5 \quad w_6$

$$\frac{\text{TF} \times \text{IDF}}{1000000}$$

10  
 $\text{TF} \times 1000000$  → high

$$\frac{\log_{10}(1000000)}{10}$$

$$TF \times \log(Idf)$$

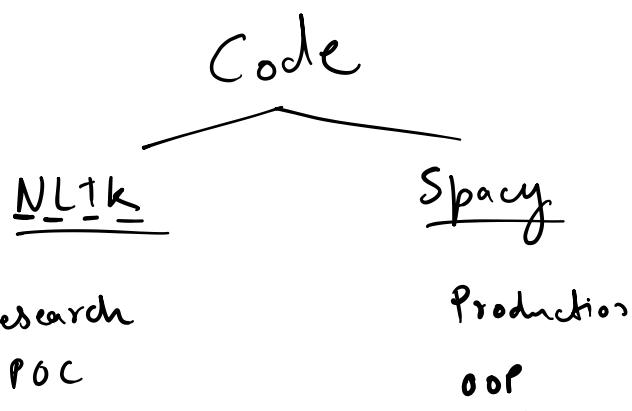
$$TF \times \log_{10} \left( \frac{N+1}{df+1} \right) + 1$$

$$\frac{10+1}{0+1} \rightarrow \infty$$

Smoothing

$$\log \frac{10}{10} = 0$$

$$\frac{1}{1000000} \dots \dots \dots$$



### Cleaning

- url remove
- punctuations — remove ! ( - ) ,
- digit .

1. 1. 1

food is good   
food is good

Remove url → Regex  
"pattern"

{ https:// . . . . .  
http:// . . . . .  
www. http . . . . .

Chaitin → use regex to find url

\S

Mapping text replaced delete  
trans = str.maketrans("aeiou", "12345"), ''  
text = "education"  
translated = text.translate(trans)  
translated

str maketrans ("aeiou", "12345"), "nea"  
String S h e i o u 1 2 3 4 5 n e a

play   
dancing   
running 

Stemming : fast

~~play~~ ~~it's~~ ~~buy~~      caring  
                        → car

lemmatization — Lemma - {  
    playing - play }  
    bookings - book  
    caring - care