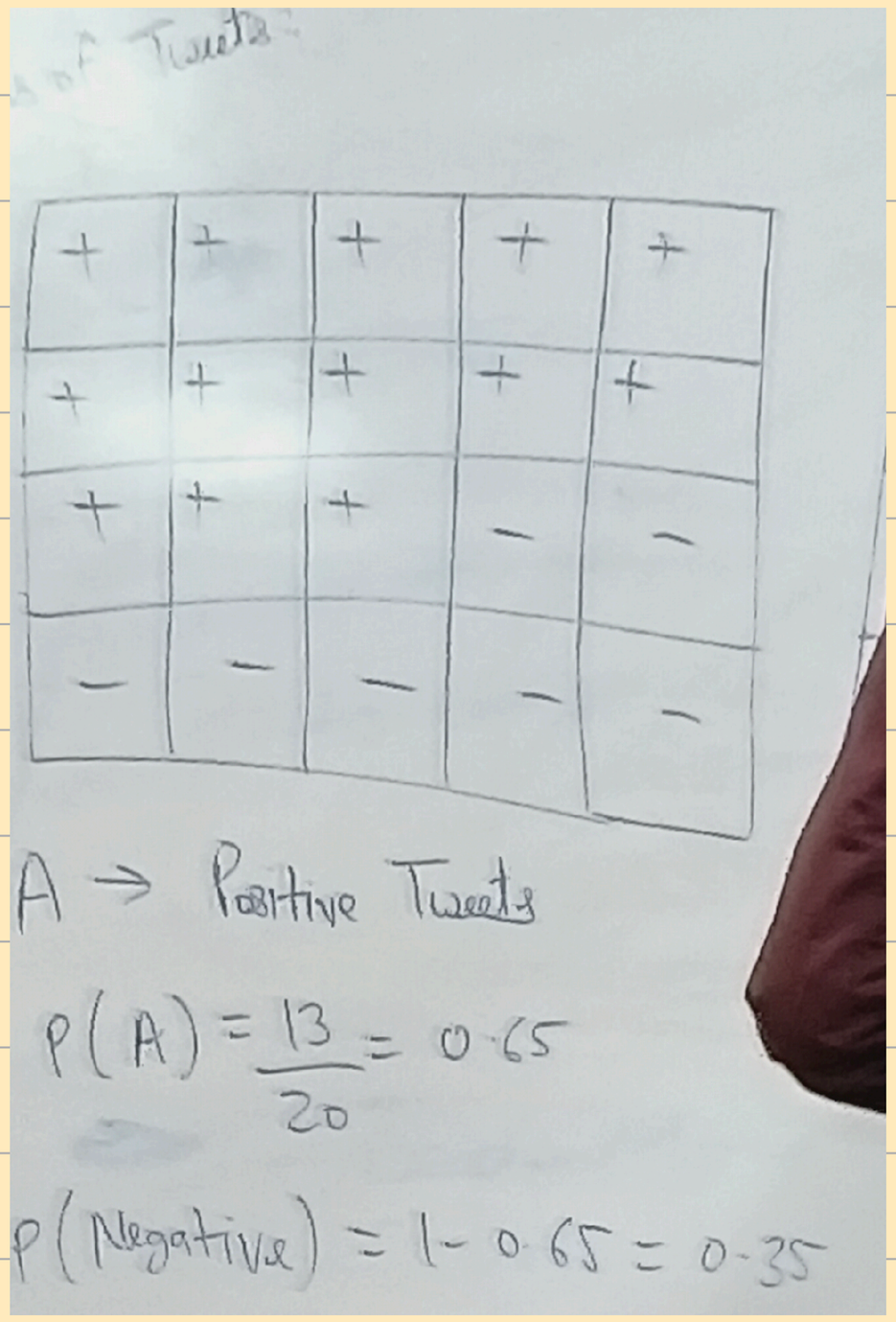
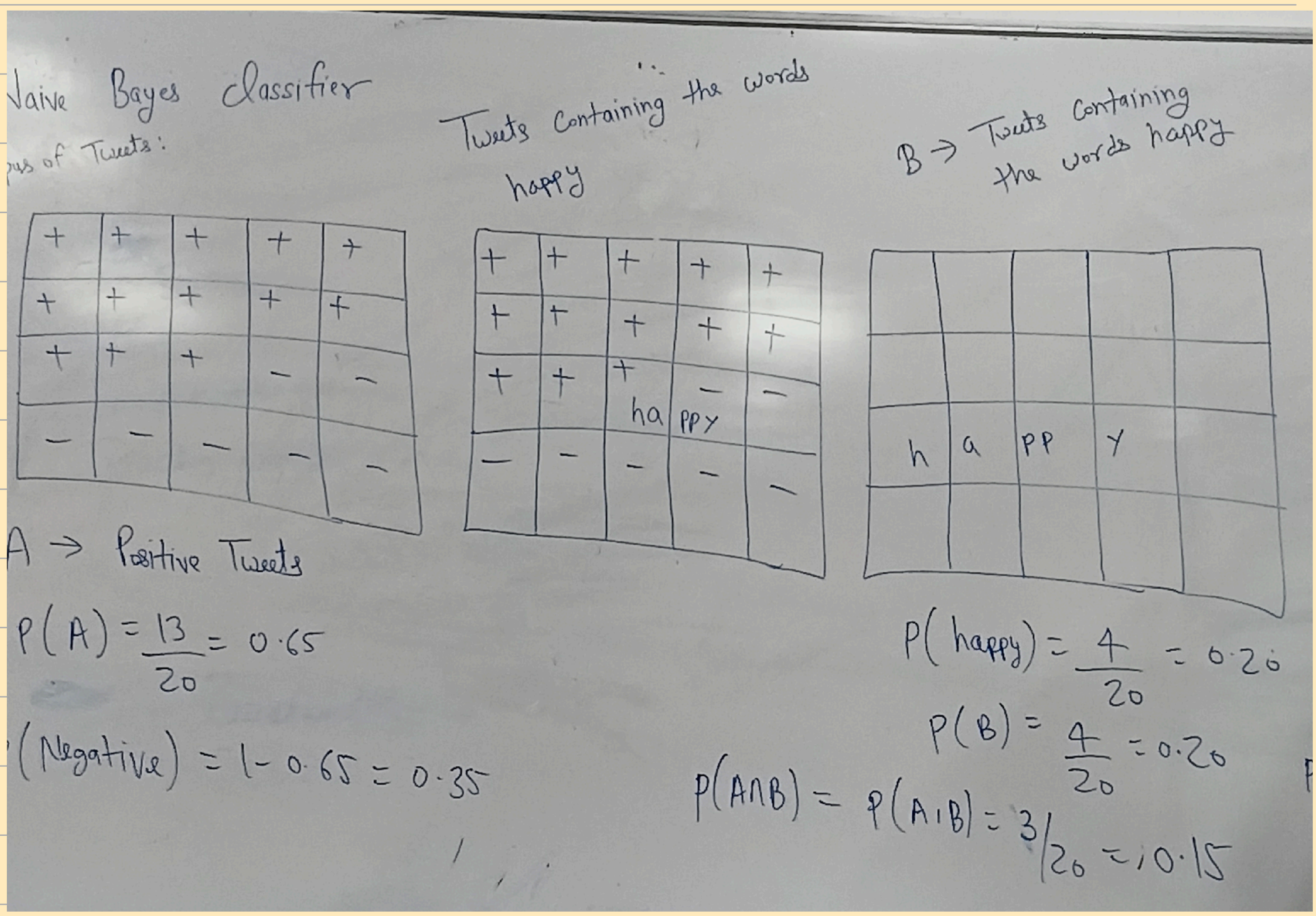
15/01/26

Naive Bayes Classifier :





s Containing the words

happy

B → Tweets Containing the words happy

+	+	+	+	
+	+	+	+	
+	+	−	−	
	ha	ppy		
−	−	−	−	

h	a	pp	y	

$$P(happy) = \frac{4}{20} = 0.20$$

$$P(B) = \frac{4}{20} = 0.20$$

$$P(A \cap B) = P(A,B) = 3/20 = 0.15$$

$$P(A/B) = P(Positive/happy) = 3/4 = 0.75$$

$$P(B/A) = P(happy/Positive) = 3/13 = 0$$

$$P(X/Y) = \frac{P(Y/X)\,P(X)}{P(Y)}$$

Corpus

I am happy, because I am learning NLP — +ve

I am happy, not sad.

I am sad, I am [redacted] learning NLP. — -ve

I am sad, not happy

$$P(A/B) = P(Positive/happy) = 3/4 = 0.75$$

$$P(B/A) = P(happy/Positive) = 3/13 = 0.231$$

Vocab	Positive	Negative	P(Wi/Positive)	P(Wi/Negative)
I	3	3	$\frac{3}{13} = 0.24$	$\frac{3}{12} = 0.24$
am	3	3	0.24	0.24
happy	2	1	0.15	0.08
because	1	0	0.08	0.0
learning	1	1	0.08	0.08
NLP	1	1	0.08	0.08
sad	1	2	0.08	0.15
not	1	1	0.08	0.15
	13	13	1	1

$$P(X/Y) = \frac{P(Y/X)\,P(X)}{P(Y)}$$

Corpus

I am happy, because I am learning NLP. +ve

I am happy, not sad.

I am sad, I am not learning NLP. -ve

I am sad, not happy

Tweet: I am happy today I am learning.
X X

I am not sad.

$$\frac{P(+ve)}{P(-ve)} \begin{pmatrix} 0.15 \\ 0.08 \end{pmatrix} \Rightarrow \text{Positive}$$

20/01/26

$$P(C/X) = \frac{P(X/C)\,P(C)}{P(X)} \Longrightarrow$$ Constant for all

C: class X: Document

P(C) : Prior probability of class

P(X/C) = Liklihood

P(C/X) = Posterior Probability

$$P(C/X) \propto P(C) \prod_{i=1}^{n} P(w_i/C)$$

Naive Independence assumption

for $X = \{w_1, w_2 \dots w_n\}$

t the class $P(X/c) = \prod_{i=1}^{n} P(w_i/c)$

Ex

Means: Presence of one word does not affect the probability of another word.

Naive: Assume that feature (words in this case) Conditionally independant given class.

Steps: . Text Preprocessing (tokenization, stop word removal)
. Feature Extraction
. Compute: Class Prior, Likelihood, Posterior
. Choose Class with maximum probability.

Schedule, Today, Project } c = 9

X = Free Money

$P(Spam/X) = \frac{P(X/Spam) \times P(Spam)}{P(X)}$

$P(Ham/X) = \frac{P(X/Ham) \times P(Ham)}{P(X)}$

P(Ham) = P(Spam) = 0.5

$P(Spam/X) \propto P(Spam) P(X/Spam)$

$\propto 0.5 \times P(Free/Spam) \times P(Money/Spam)$

$\propto 0.5 \times \frac{2+1}{6+9} \times \frac{1+1}{6+9}$

$\propto 0.0133$

$$P(X/c) = \prod_{i=1}^{n} P(w_i/c)$$

Means: Presence of one word does not affect the probability of another word.

Likelihood with Laplacian Smoothing.

$$P(w/c) = \frac{\text{Count}(w, c) + 1}{\text{total words in } c + |V|}$$

$$P(\text{Ham}/X) \propto P(\text{Ham}) \times P(X/\text{Ham})$$

$$\propto 0.5 \times P(\text{Free}/\text{Ham}) \times P(\text{Money}/\text{Ham})$$

$$\propto 0.5 \times \frac{0+1}{6+9} \times \frac{0+1}{6+9} = 0.0022$$

Ex

Document	Text	Class
D_1	Free Monthly Offer	Spam
D_2	Free Prize Money	Spam
D_3	Meeting Schedule Today	Ham
D_4	Project Meeting Today	Ham

Test Document : Free Money

$P(\text{Spam}/X) > P(\text{Ham}/X)$

Class ⇒ Spam

22/01/26

Reviews	Text	Sentiment
R_1	good movie enjoyable.	Positive
R_2	good acting good story.	Positive
R_3	bad movie boring.	Negative
R_4	bad acting boring movie.	Negative

Test: Good Movie Boring

$Vocab = \{$ good, movie, enjoyable, acting, story, bad, boring $\}$

$|V| = 7$

$$P(C/X) = \frac{P(X/C)\,P(C)}{P(X)}$$

$P(C/X)$: Posterior; $P(X/C)$: likelihood; $P(C)$: Prior $\in$ Prior; $P(X)$: same for all class

$X = [w_1, w_2 \ldots w_n]$

$P(w_i/c)$

$\log(?)$

$\log P(X/C) = \log \prod_{i=1}^{n} P(w_i/c)$

$P(C/X) \propto P(C) \cdot P(X/C)$

$\log P(C/X) \propto \log\left(P(C) \prod_{i=1}^{n} P(w_i/c)\right) = \log[P(C)] + \log \prod_{i=1}^{n} P(w_i/c)$

$\log P(\text{Positive}/\text{Good, Movie, Boring})$

$\log P(\text{Negative}/\text{Good Movie, Boring})$

Test: Good Movie Boring

$Vocab = \{$ good, movie, enjoyable, acting, story, bad, boring $\}$

$|V| = 7$