Agenda

- What is Naive Bayes?
- Naive Bayes and Machine Learning
- Why do we need Naive Bayes?
- Understanding Naive Bayes Classifier
- Advantages of Naive Bayes Classifier
- Demo - Text Classification using Naive Bayes

Prereq

Conditional prob revision

brief → Intro NLP

LET US CONSIDER THE FOLLOWING EXAMPLE OF TOSSING TWO COINS

Here, the sample space is:

{HH, HT, TH, TT}

1. P(Getting two heads) = 1/4
2. P(At least one tail) = 3/4
3. P(Second coin being head given first coin is tail) = 1/2
4. P(Getting two heads given first coin is a head) = 1/2

Bayes' Theorem gives the conditional probability of an event A given another event B has occurred

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where:

P(A|B) = Conditional Probability of A given B

P(B|A) = Conditional Probability of B given A

P(A) = Probability of event A

P(B) = Probability of event A

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A/B)\,P(B) = P(A \cap B)$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(B/A)\,P(A) = P(A \cap B)$$

Bayes

$$\Rightarrow P(A/B)\,P(B) = P(B/A)\,P(A)$$

ACTION

P

IN THIS SAMPLE SPACE, LET A BE THE EVENT THAT SECOND COIN IS HEAD AND B BE THE EVENT THAT FIRST COIN IS TAIL

A: 2nd Head
B: 1st Tail

2/4

In the sample space:

{HH, HT, TH, TT}

P(Second coin being head given first coin is tail)

= P(A|B)

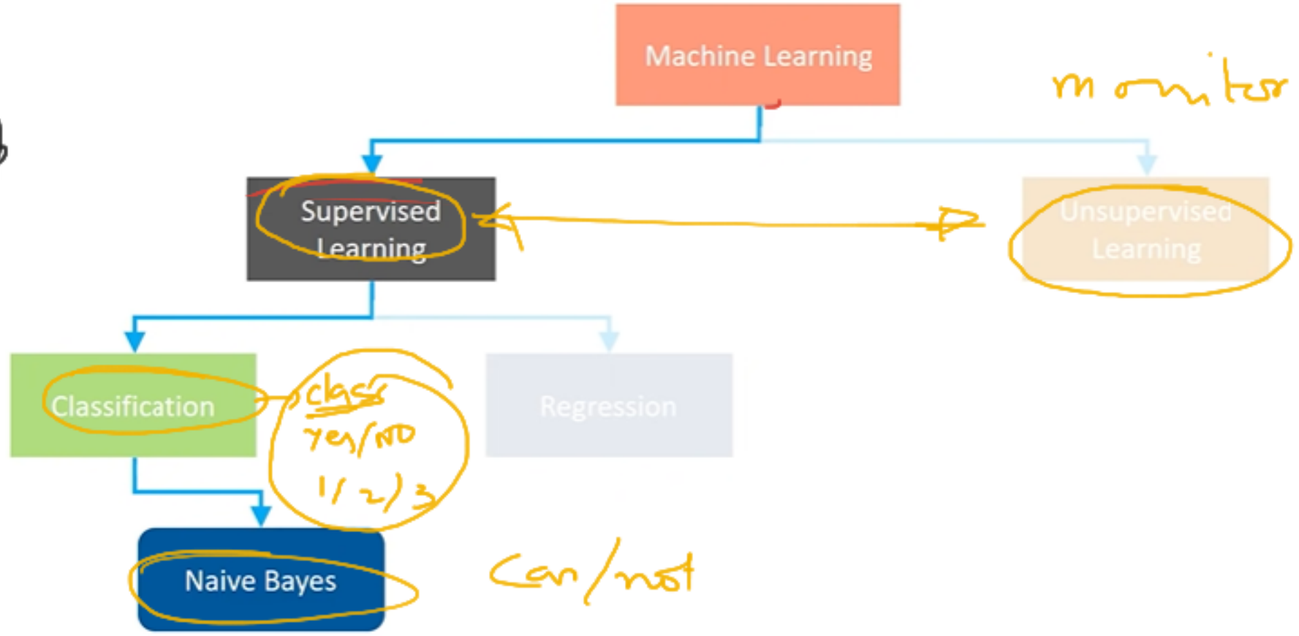
= [P(B|A) * P(A)] / P(B)

= [P(First coin being tail given second coin is head) * P(Second coin being head)] / P(First coin being tail)

= [(1/2) * (1/2)] / (1/2)

= 1/2 = 0.5

P(1st tail / 2nd H)



Where is Naive Bayes used?

Face Recognition

Weather Prediction

Medical Diagnosis

News Classification

P(Play / 2/50 , Punish , Bunk) situation = > 1

P(NP / ⟵⟶)

< 1 — Less / More — < 1

Naive Bayes — indep ? decision

$$P(Y \mid X) = P(y)P(x_1|y)P(x_2|y,x_1)\ldots P(x_n|y,x_1,\ldots,x_{n-1})$$

$x_1 x_2 x_3$

$P(Y/X) = P(Y)\, P(x_1/y)\, P(x_2/y, x_1)$ Int

Naive (Simple)

4/9 × × ×

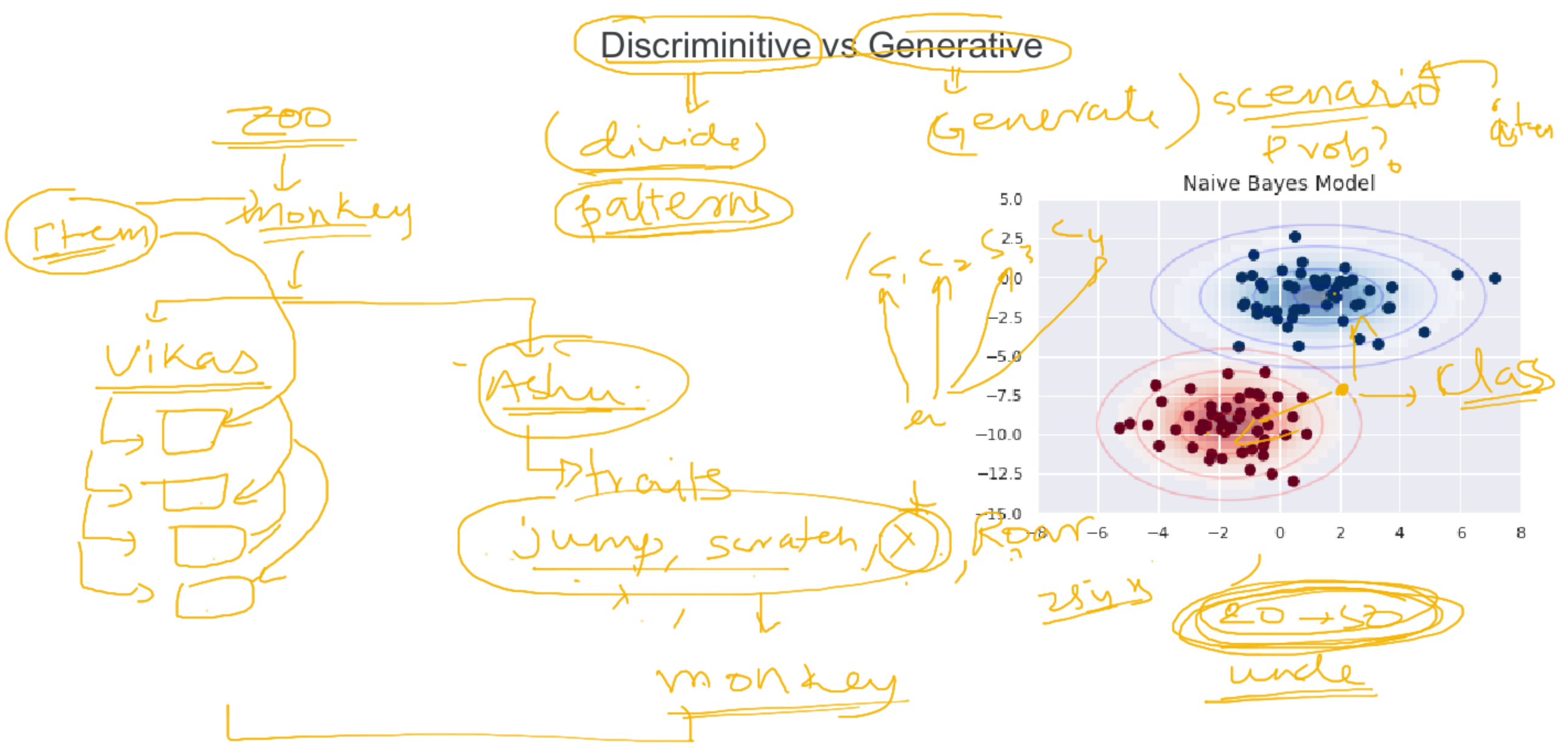
2/50 — Punish — Bunk

Real

X_1 X_2 X_3

independ each oth

Discriminitive vs Generative



P( )

To predict whether a person will purchase a product on a specific combination of **Day**, **Discount** and **Free Delivery** using **Naive Bayes Classifier**

Day	Discount	Free Delivery
Weekdays	Yes	Yes
weekend		
HOLIDAY	No	No

Probabilities → Buy / Don't Buy

Generate?.

Dataset

We have a small sample dataset of 30 rows for our demo

	A	B	C	D
1	Day	Discount	Free Delivery	Purchase
2	Weekday	Yes	Yes	Yes
3	Weekday	Yes	Yes	Yes
4	Weekday	No	No	No
5	Holiday	Yes	Yes	Yes
6	Weekend	Yes	Yes	Yes
7	Holiday	No	No	No
8	Weekend	Yes	No	Yes
9	Weekday	Yes	Yes	Yes
10	Weekend	Yes	Yes	Yes
11	Holiday	Yes	Yes	Yes
12	Holiday	No	Yes	Yes
13	Holiday	No	No	No
14	Weekend	Yes	Yes	Yes
15	Holiday	Yes	Yes	Yes

Naive_Bayes_Dataset

Based on this dataset containing three input types of ***Day***, ***Discount*** and ***Free Delivery***, we will populate frequency tables for each attribute

Pivot tables

Pic 1

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Pic 2

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

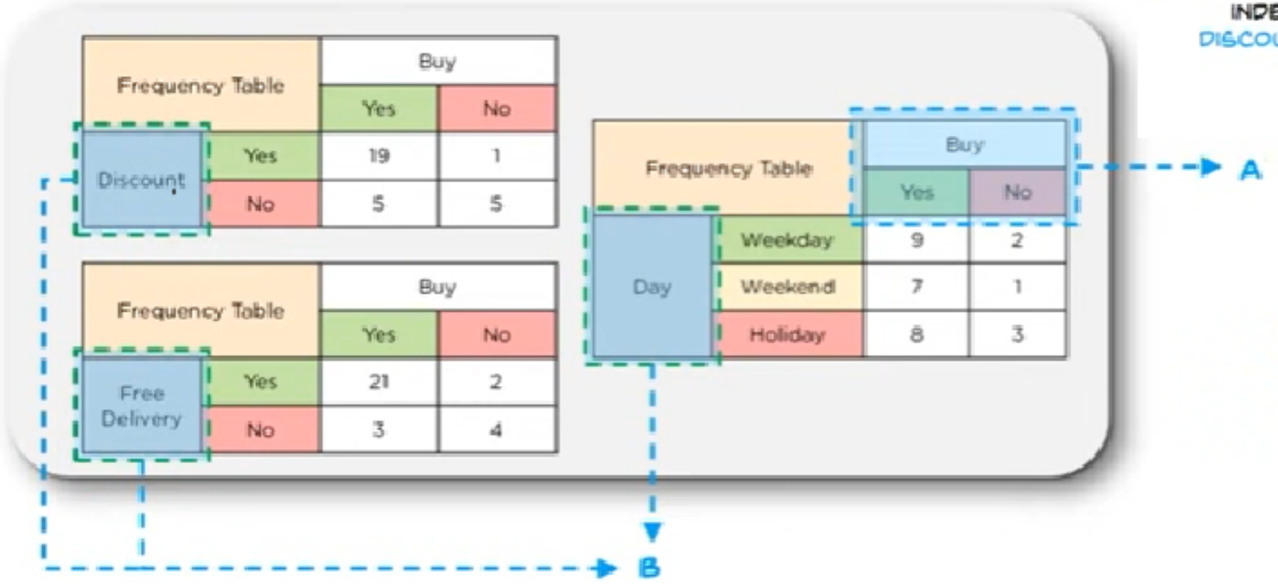
step

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Pic 3

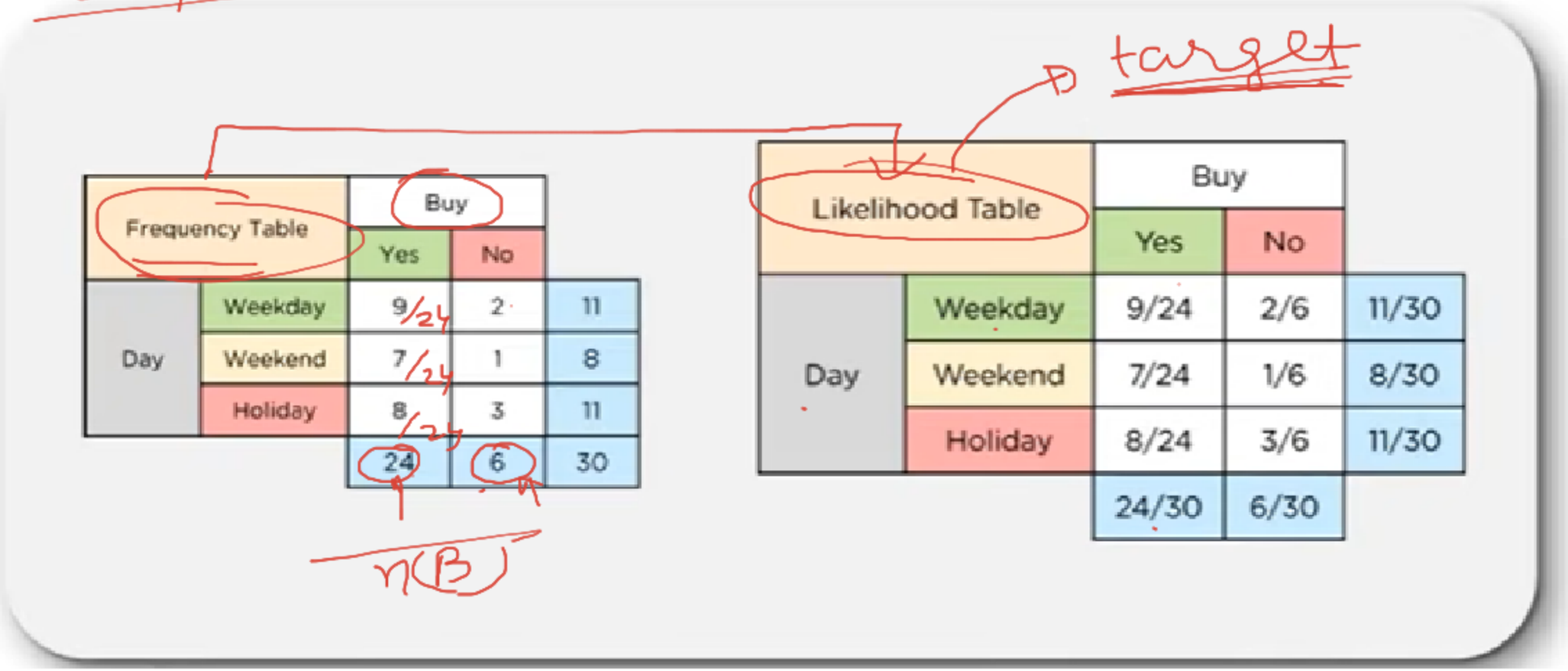
Based on this dataset containing three input types of ***Day***, ***Discount*** and ***Free Delivery***, we will populate frequency tables for each attribute

FOR OUR BAYES THEOREM, LET THE EVENT BUY BE A AND THE INDEPENDENT VARIABLES, DISCOUNT, FREE DELIVERY AND DAY BE B



Now let us calculate the Likelihood table for one of the variable, *Day* which includes ***Weekday***, ***Weekend*** and ***Holiday***

Step 1



Frequency Table

Day	Buy = Yes	Buy = No	
Weekday	9/24	2	11
Weekend	7/24	1	8
Holiday	8/24	3	11
	24	6	30

n(B)

Likelihood Table

Day	Buy = Yes	Buy = No	
Weekday	9/24	2/6	11/30
Weekend	7/24	1/6	8/30
Holiday	8/24	3/6	11/30
	24/30	6/30	

target

P(B) = P(Weekday)
= 11/30 = 0.37

P(A) = P(No Buy)
= 6/30 = 0.2

P(B|A)
= P(Weekday | No Buy)
= 2/6 = 0.33

P(A|B)

Based on this likelihood table, we will calculate conditional probabilities as below

Frequency Table		Buy		
		Yes	No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

P(B) = P(Weekday) = 11/30 = 0.367

P(A) = P(No Buy) = 6/30 = 0.2

P(B|A) = P(Weekday | No Buy) = 2/6 = 0.33

Prod NB

P (A|B) = P (No Buy | Weekday)

= P(Weekday| No Buy) * P(No Buy) / P(Weekday)

= (0.33 * 0.2) / 0.367 = 0.179

2/11

LET US USE THESE 3 LIKELIHOOD TABLES TO CALCULATE WHETHER A CUSTOMER WILL PURCHASE A PRODUCT ON A SPECIFIC COMBINATION OF DAY, DISCOUNT AND FREE DELIVERY OR NOT

HERE, LET US TAKE A COMBINATION OF THESE FACTORS:

- DAY = HOLIDAY
- DISCOUNT = YES
- FREE DELIVERY = YES

Likelihood Tables

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Table		Buy		
		Yes	No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Table		Buy		
		Yes	No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

Calculating Conditional Probability of purchase on the following combination of day, discount and free delivery:

Where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let A = No Buy

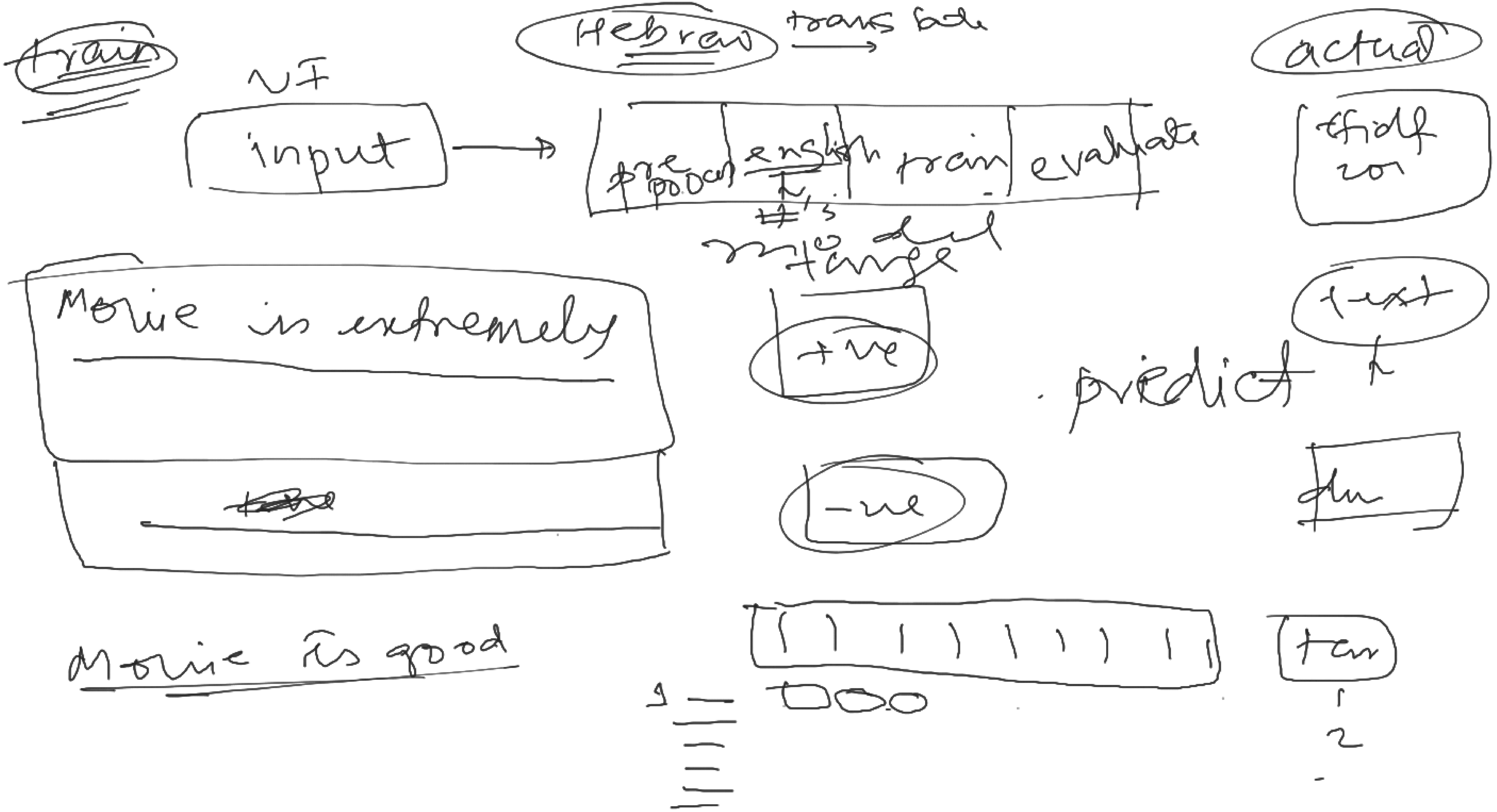
P(A|B) = P(No Buy | Discount = Yes, Free Delivery = Yes, Day = Holiday)

$$= \frac{P(\text{Discount = Yes} \mid \text{No}) * P(\text{Free Delivery = Yes} \mid \text{No}) * P(\text{Day = Holiday} \mid \text{No}) * P(\text{No Buy})}{P(\text{Discount=Yes}) * P(\text{Free Delivery=Yes}) * P(\text{Day=Holiday})}$$

$$= \frac{(1/6) * (2/6) * (3/6) * (6/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.178$$

$$P(Y=1|X) = \frac{P(y=1)P(x_1|y=1)....P(x_n|y=1)}{P(X)}$$


train
Hebrew
actual
input
pre process
english
train
evaluate
tfidf
target
Movie is extremely
+ve
-ve
text
predict
Movie is good

SUM OF PROBABILITIES
= 0.986 + 0.178 = 1.164

LIKELIHOOD OF PURCHASE
= 0.986 / 1.164 = 84.71 %

LIKELIHOOD OF NO PURCHASE
= 0.178 / 1.164 = 15.29 %

PROBABILITY OF PURCHASE = 0.986
PROBABILITY OF NO PURCHASE = 0.178

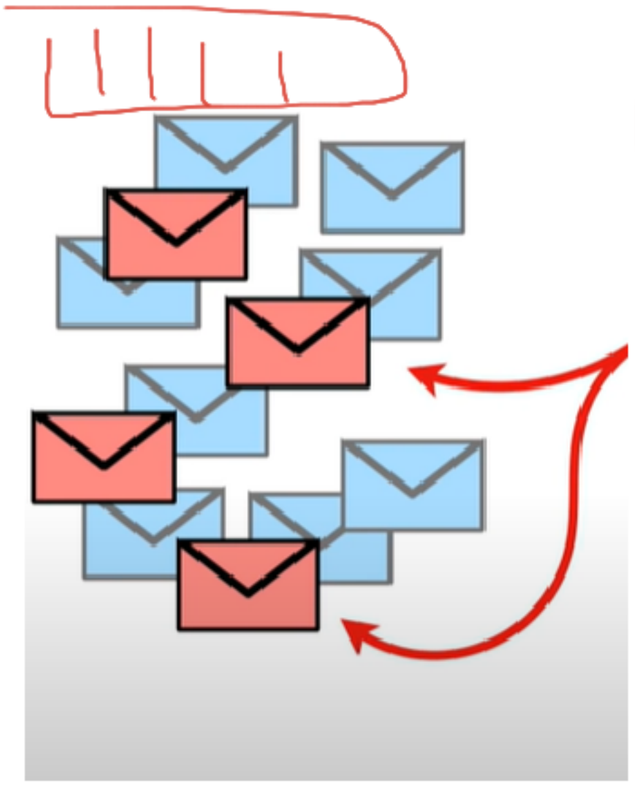
P( )

grammers

Spam

ham

count-tfidf



R_1 sky is blue

R_2 Rose is Red blue Red

Preprocessing

clean

sky is blue Rose Red

R_1	1	1	1	0	0
R_2	0	1	1	1	2

NLP → punc

stop

computer → vectorized

english → M/c's

→ integer →

data

	W_1	W_2	W_3	W_4	W_5	target
R_1						
R_2						
R_3						

msg = Dear Friend Lunch money

LT

p(**Dear** | **N**) = 0.47
p(**Friend** | **N**) = 0.29
p(**Lunch** | **N**) = 0.18
p(**Money** | **N**) = 0.06

likelihood

p(**Dear** | **N**) = 0.47
p(**Friend** | **N**) = 0.29
p(**Lunch** | **N**) = 0.18
p(**Money** | **N**) = 0.06

p(**Dear** | **S**) = 0.29
p(**Friend** | **S**) = 0.14
p(**Lunch** | **S**) = 0.00
p(**Money** | **S**) = 0.57

p(**Dear** | **S**) = 0.29
p(**Friend** | **S**) = 0.14
p(**Lunch** | **S**) = 0
p(**Money** | **S**) = 0.57

Dear Friend Lunch Money

predict msg = **Dear Friend**

$p(\text{N}) \times p(\text{Dear} \mid \text{N}) \times p(\text{Friend} \mid \text{N}) = 0.09$

$p(\text{S}) \times p(\text{Dear} \mid \text{S}) \times p(\text{Friend} \mid \text{S}) = 0.01$

Lunch Money Money Money Money

$p(\text{Dear} \mid \text{N}) = 0.47$
$p(\text{Friend} \mid \text{N}) = 0.29$
$p(\text{Lunch} \mid \text{N}) = 0.18$
$p(\text{Money} \mid \text{N}) = 0.06$

$p(\text{N}) = 0.67$

…and that means we will always classify the messages with **Lunch** in them as **normal**, no matter how how many times we see the word **Money**.

$p(\text{N}) \times p(\text{Lunch} \mid \text{N}) \times p(\text{Money} \mid \text{N})^4 = 0.000002$

$p(\text{S}) \times p(\text{Lunch} \mid \text{S}) \times p(\text{Money} \mid \text{S})^4 = 0$

$p(\text{Dear} \mid \text{S}) = 0.29$
$p(\text{Friend} \mid \text{S}) = 0.14$
$p(\text{Lunch} \mid \text{S}) = 0.00$
$p(\text{Money} \mid \text{S}) = 0.57$

$p(\text{S}) = 0.33$

Lunch Money Money Money Money

p(Dear | N) = 0.43
p(Friend | N) = 0.29
p(Lunch | N) = 0.19
p(Money | N) = 0.10

p(N) = 0.67

p(Dear | S) = 0.27
p(Friend | S) = 0.18
p(Lunch | S) = 0.09
p(Money | S) = 0.45

p(S) = 0.33

And since the value for **spam** is greater than the one for a **normal message**...

$$p(\mathbf{N}) \times p(\mathbf{Lunch} \mid \mathbf{N}) \times p(\mathbf{Money} \mid \mathbf{N})^4 = 0.00001$$

$$p(\mathbf{S}) \times p(\mathbf{Lunch} \mid \mathbf{S}) \times p(\mathbf{Money} \mid \mathbf{S})^4 = 0.00122$$


5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100