$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|}$$

Training	Doc	Words	Class
Training	1.	Chinese beijing Chine	c
	2.	Chinese Chinese Shanghai	c
	3.	Chinese Macao	c
	4.	Tokyo Japan chinese.	j
Test	5.	Chinese Chinese Chines Tokyo Japan.	?

Priors

$P(c) = 3/4$

$P(j) = 1/4$

Conditional Probabilities

$$P(\text{Chinese}|c) = \frac{5+1}{5+6} = \frac{6}{14} = \frac{3}{7}$$

$$P(\text{Tokyo}|c) = (0+1)/(8+6) = \frac{1}{14}$$

$$P(\text{Japan}|c) = (0+1)/(8+6) = \frac{1}{14}$$

$$P(\text{Chinese}|j) = \frac{(1+1)}{(3+6)} = \frac{2}{9}$$

$$P(\text{Tokyo}|j) = \frac{1+1}{3+6} = \frac{2}{9}$$

$$P(\text{Japan}|j) = \frac{(1+1)}{3+6} = \frac{2}{9}$$

Test, $P(C|d5) \propto \frac{3}{4} \times \left(\frac{3}{7}\right)^3 \times \frac{1}{14} \frac{1}{14} = 0.0003$

$P(j|ds) \propto \frac{1}{4} \cdot \left(\frac{2}{9}\right)^3 \times \frac{2}{9} \times \frac{2}{9} \approx 0.0001.$

Next Example

	Cat	Documents	
Training	-	just plain boring	3/5
	-	entirely predictable and lacks energy	3/5
	-	no surprises and very few laughs	3/5
	+	very powerful	2/5
	+	the most fun film of the summer	2/5
Test	?	predictable with no fun.	

#

	Weather	Play
1	Sunny	No 1
	Overcast	Yes
	Rainy	Yes
2 ✓	Sunny	Yes ✓
3	Sunny	Yes ✓
	Overcast	Yes
	Rainy	No 2
	Rainy	No 3
4	Sunny	Yes ✓
	Rainy	Yes
5	Sunny	No 4
	Overcast	Yes
	Overcast	Yes
	Rainy	No 5

Q. what is the probability of $P(Yes \mid Sunny)$

$$P(Yes \mid sunny) = \frac{P(sunny \mid yes) \cdot P(yes)}{P(sunny)}$$

$$= \frac{3/9 \cdot 9/14}{5/14} = \frac{3}{5}$$

$$P(No \mid sunny) = \frac{P(Sunny \mid No) \cdot P(No)}{P(sunny)}$$

$$= \frac{2/5 \cdot 5/14}{5/14} = \frac{2}{5}$$

Next Example

	Cat	Documents
Training	-	just plain boring (1) (2) (3)
	-	entirely predictable and lacks energy (4) (5) (6) (7) (8)
	-	no surprises and very few laughs (9) (10) (11) (12) (13) (14)
	+	very powerful (15) (16)
	+	the most fun film of the summer 17 18 19 20 21 22 23
Test	?	predictable with no fun.

Q. Find ? i.e. the category of the above test document.

$P(+) = \frac{2}{5}$

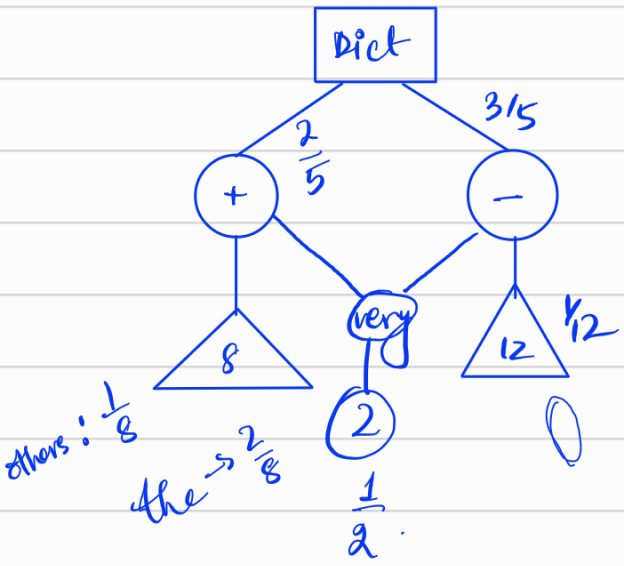
$P(-) = \frac{3}{5}$

$P(\text{predictable}) = \frac{1}{23}$

$P(\text{with}) = 0$

$P(\text{no}) = 1/23$

$P(\text{fun}) = 1/23$



$$P(\text{Predictable} \mid -) = \frac{1+1}{14+20}$$

$$= \frac{2}{34}$$

$$= \frac{1}{17}$$

$$P(\text{Predictable} \mid +) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$P(\text{no} \mid -) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$$

$$P(\text{no} \mid +) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$P(\text{fun} \mid -) = \frac{0+1}{14+20} = \frac{1}{34}$$

$$P(\text{fun}|+) = \frac{1+1}{9+20} = \frac{2}{29}$$

Scoring

$$P(-)\ P(D|-) = \frac{3}{5} \times \frac{2 \times 2}{34^3}$$

$$\simeq 6 \times 10^{-5}$$

$$P(+)\ P(D|+) = \frac{2}{5} \times \frac{2}{29^3}$$

$$\simeq 3.2 \times 10^{-5}$$

As, $P(-)\ P(D|-) > P(+)\ P(D|+)$

So, we can say that the test document belongs to the negative category.