prize

lottery, coupon, bonus

play → ply → play

playing plays → play

improving improve improved

Spam Mail Detection

The movie is good and scray. →

1.) Bag of Wovd

2.) TF-IDF

1.) Text processing is necessary. → D1

2.) Text processing is necessary and important. → D2

3.) " " " easy. → D3

{D1, D2, D3} → corpus

Vocabulary → Unique words in a corpus.

'Text' 'text' lower()

	Text	Processing	is	necessary	and	important	easy
D1	1	1	1	1	0	0	0
D2	1	1	2	1	1	1	0
D3	1	1	1	0	0	0	1

TF – IDF

TF → Term Frequency

IDF → Inverse Document Frequency

$$TF = \frac{\text{No. of times term 't' appears in a doc.}}{\text{No. of terms in a document}}$$

$TF(Text) = \frac{1}{4}$

$TF(Pro) = 1/4$

$TF(is) = 1/4$

$TF(nec) = 1/4$

$TF(and) = 0$

$TF(imp) = 0$

$TF(easy) = 0$

Terms	TF(D1)	TF(D2)	TF(D3)
Text	1/4	1/6	1/4
Pro	1/4	1/6	1/4
is	1/4	2/6	1/4
nec	1/4	1/6	0
and	0	1/6	0
imp.	0	1/6	0
easy	0	0	1/4

$$IDF = \log \frac{\text{no. of documents}}{\text{no. of document with term 't'}}$$

$$= \log \frac{3}{3} \Rightarrow \log(1) = 0$$

$$IDF(\text{necessary}) = \log \frac{3}{2}$$

Terms	D_1	D_2	D_3	IDF	TfIDF(D_1)	TFI(D_2)	TF (D_3)
Text	1	1	1	0	0	0	0
Proc.	1	1	1	0	0	0	0
is	1	2	1	0	0	0	0
nece.	1	1	0	$\log(3/2)$	$1/4 \times \log(3/2)$	$1/6 \times \log(3/2)$	0
and	0	1	0	$\log(3)$	0	$1/6 \times \log(3)$	0
imp.	0	1	0	$\log(3)$	0	$1/6 \times \log(3)$	0
easy	0	0	1	$\log(3)$	0	0	$1/4 \times \log(3)$

Naive Bayes Classifier

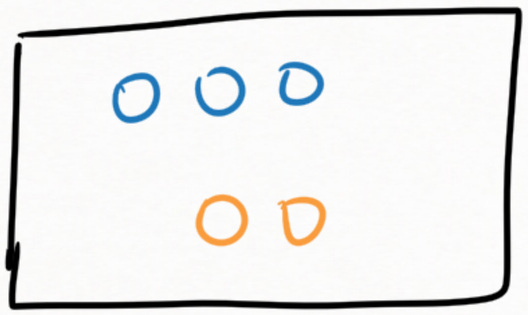
Independent Event → What is probability of getting head after tossing coin

Dependent "

$P(H) = 1/2$, $P(T) = 1/2$

$P(1) = 1/6$

$P(2) = 1/6$



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

$$P(O) = \frac{2}{4} = \frac{1}{2}$$

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

-(i) Conditional Probability

$$A \cap B = B \cap A \quad - (ii)$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \quad - (iii)$$

$$P(B \cap A) = P(B/A) \cdot P(A)$$

Likelihood

Prior

$$P(A/B) = \frac{P(A) \cdot P(B/A)}{P(B)}$$

Probability

Posterior

Probability

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

$$P(Yes) = \frac{9}{14}$$

$$P(No) = \frac{5}{14}$$

Outlook

	Yes	No	P(Yes)	P(No)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5

Temperature

	Yes	No	P(Y)	P(N)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5

$P(\text{Yes} / \text{Rainy}, \text{Cool}) = ?$

$$P(\underbrace{\text{Yes}}_{A}/\overbrace{\text{Rainy, Cool}}^{B}) = P(\text{Yes}) * P(\text{Rainy}/\text{Yes}) * P(\text{Cool}/\text{Yes})$$

$$= \frac{9}{14} * \frac{3}{9} * \frac{3}{9}$$

$$= 0.047$$

$$P(\text{No}/\text{Rainy, Cool}) = P(\text{No}) * P(\text{Rainy}/\text{No}) * P(\text{Cool}/\text{No})$$

$$= \frac{5}{14} \times \frac{2}{5} \times \frac{1}{5} = 0.028$$

$$P(Yes/R,C) = \frac{0.047}{0.047 + 0.028}$$

$$= 0.62 \Rightarrow 62\%$$

$$P(No/R,C) = 1 - 0.62$$

$$= 0.38 = 38\%$$

$P(Sunny, Hot) = ?$