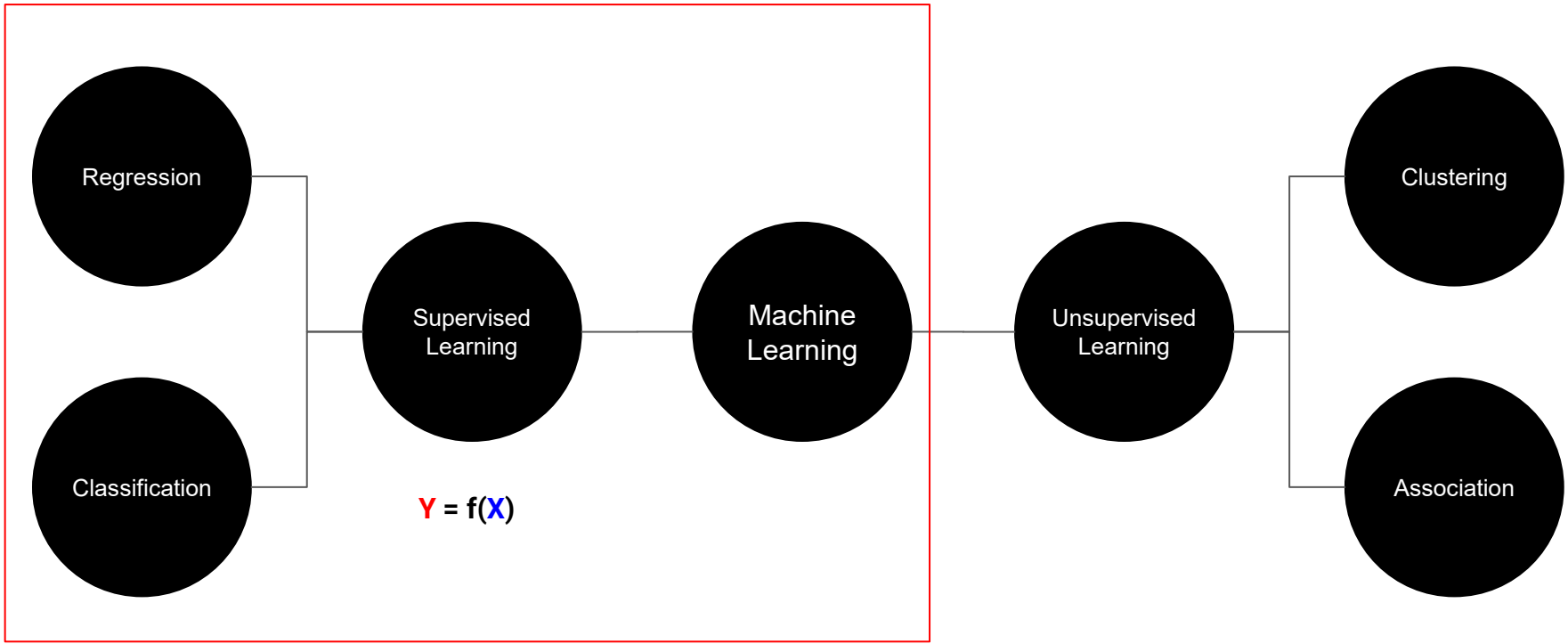
Data Prediction Model and Machine Learning

Online course #5

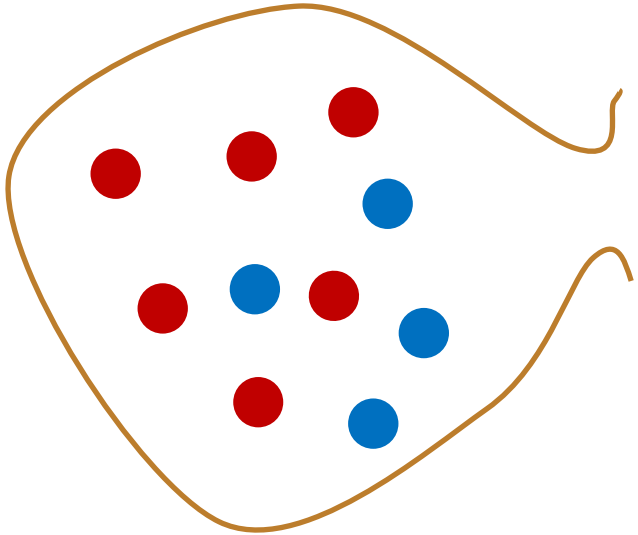
Naive Bayes


Regression
Classification
Supervised Learning
Machine Learning
Unsupervised Learning
Clustering
Association
Y = f(X)

Probability ?

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur, or how likely it is that a proposition is true.

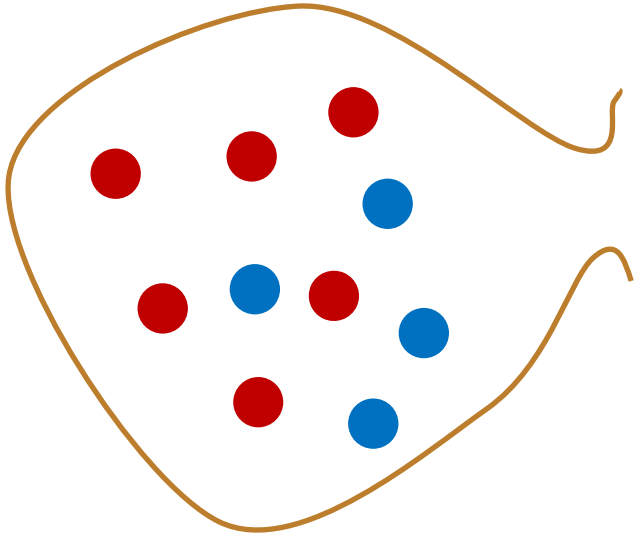
Probability ?



Probability of drawing a red ball
= P(red) = 6/10 = 60%

Probability of drawing a blue ball
= P(blue) = 4/10 = 40%

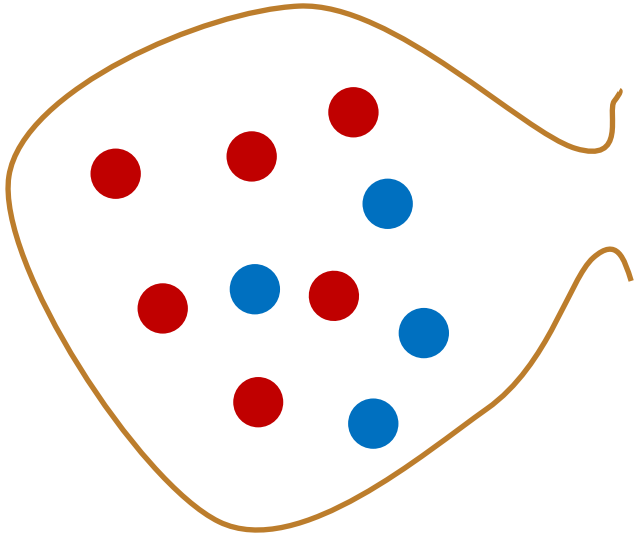
Joint probability



Let’s draw two balls!

1. With replacement
2. Without replacement

Joint probability



Let’s draw two balls!

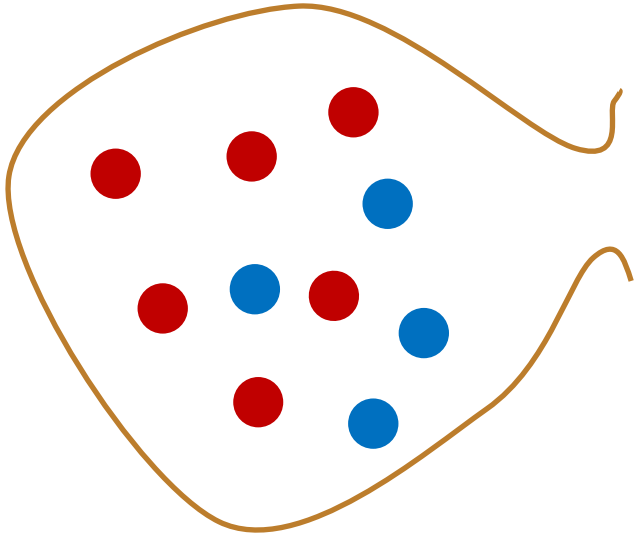
1. With replacement

Probability of drawing a red ball followed by a blue ball?

$$P(\text{red} \cap \text{blue}) = P(\text{red}) \times P(\text{blue})$$

Independent!

Joint probability



Let’s draw two balls!

2. Without replacement

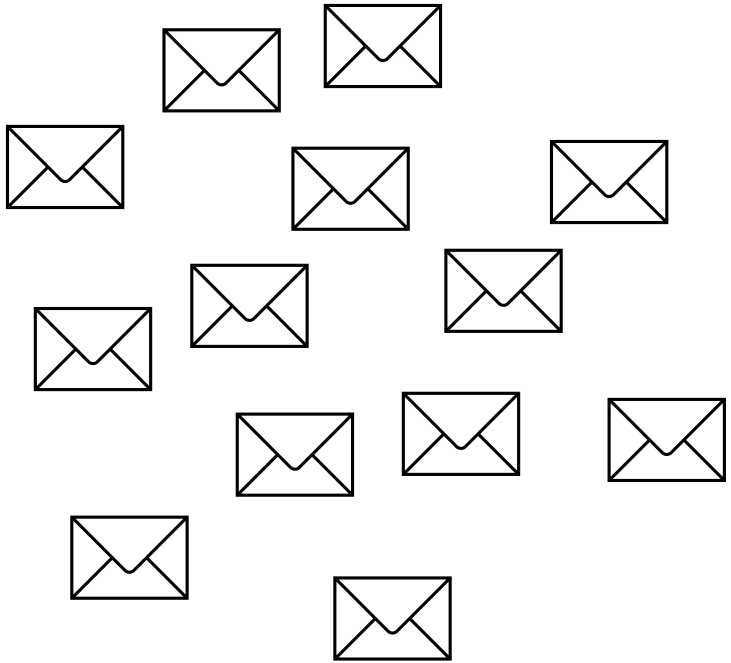
Probability of drawing a red ball followed by a blue ball?

$$P(\text{red} \cap \text{blue}) = P(\text{red}) \times P(\text{blue} \mid \text{red})$$

Conditional Probability

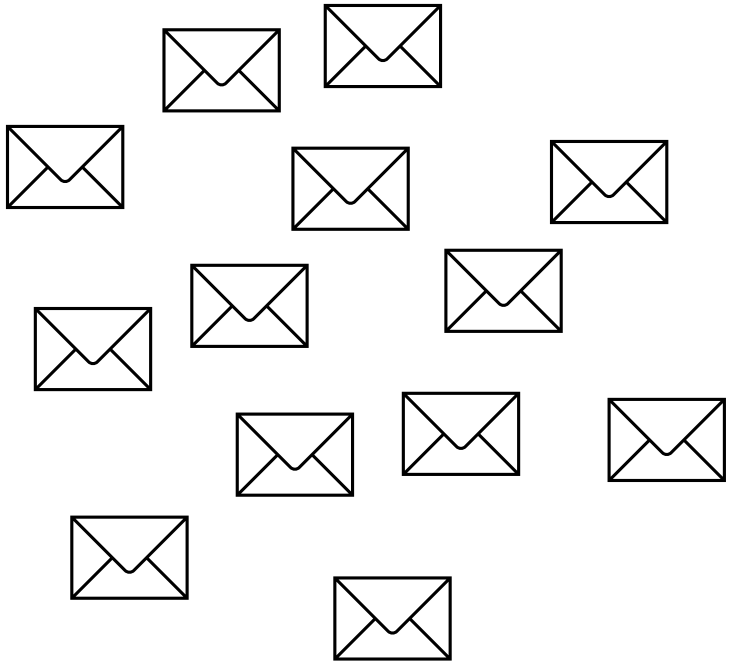
Naïve Bayes Classifier

You’ve got mails!



You’ve got mails!

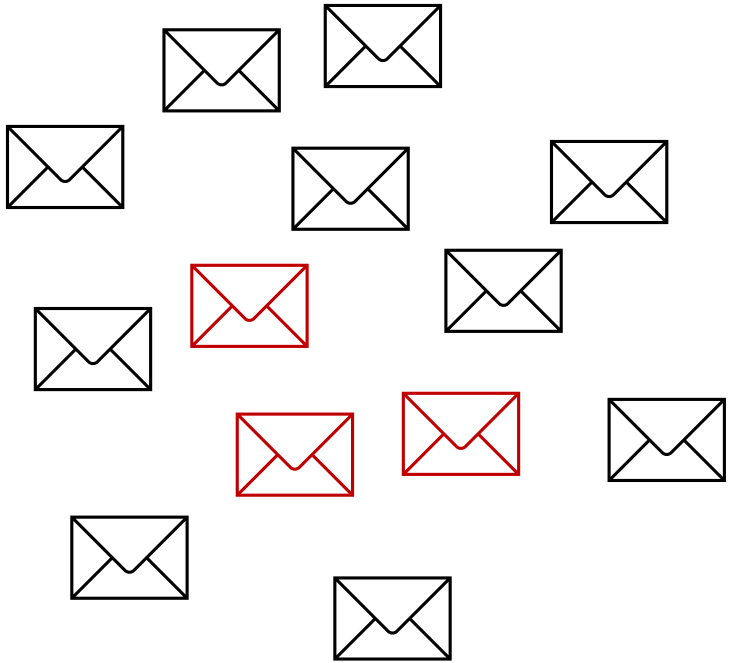
Normal mail
(**Ham**)



Spam mail
(**Spam**)

You’ve got mails!

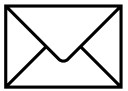
Normal mail
(**Ham**)



Spam mail
(**Spam**)

You’ve got mails!

Normal mail
(**Ham**)



Dear Chang,

How have you been?
I’m doing really well.
I’ve just had a ramen for lunch, and it reminds me of the memory when we were in Japan.
Hope you are fine too.
Talk to you very soon.

Best regards,
Your friend, Silvia

Spam mail
(**Spam**)

You’ve got mails!

Normal mail
(**Ham**)

Hello,

My name is Ms. Reem Ebrahim Al-Hashimi, I am the "Minister of state and Petroleum" also "Minister of State for International Cooperation" in UAE. I write to you on behalf of my other "three (3) colleagues" who has approved me to solicit for your "partnership in claiming of {us$47=Million}“ from a Financial Home in Cambodia on their behalf and for our "Mutual Benefits".

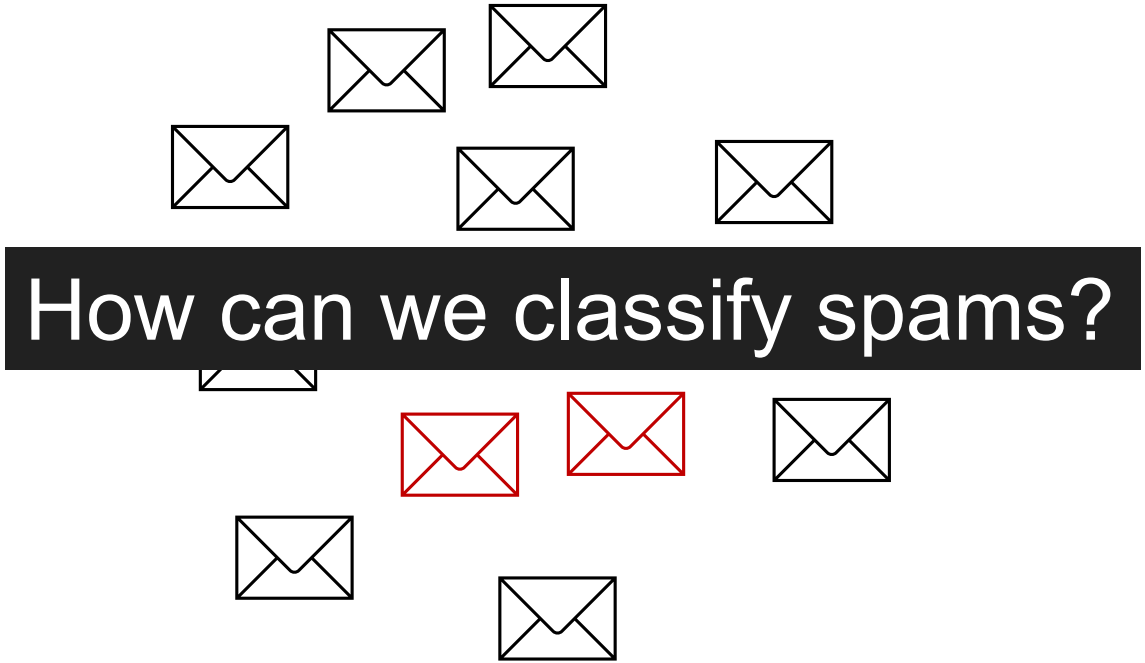
The Fund {us$47=Million} is our share from the (Over-invoiced) Oil/Gas deal with Cambodian/Vietnam Government within 2013/2014, however, We don't want our government to know about the fund. If this proposal interests you, let me know, by sending me an email and I will send to you detailed information on how this business would be successfully transacted. Be informed that nobody knows about the secret of this fund except us, and we know how to carry out the entire transaction. So I am compelled to ask, that you will stand on our behalf and receive this fund into any account that is solely controlled by you.

We will compensate you with 30% of the total amount involved as gratification for being our partner in this transaction. Reply to my private email as stated: honreemebrahimal-hashimi@yandex.com

Regards,
Ms. Reem Ebrahim Al-Hashimi.

Spam mail
(**Spam**)

You’ve got mails!



We classify the spams based on our **experience**

Hello,

My name is Ms. Reem Ebrahim Al-Hashimi, I am the "Minister of state and Petroleum" also "Minister of State for International Cooperation" in UAE. I write to you on behalf of my other "three (3) colleagues" who has approved me to solicit for your "partnership in claiming of {us$47=Million}“ from a Financial Home in Cambodia on their behalf and for our "Mutual Benefits".

The Fund {us$47=Million} is our share from the (Over-invoiced) Oil/Gas deal with Cambodian/Vietnam Government within 2013/2014, however, We don't want our government to know about the fund. If this proposal interests you, let me know, by sending me an email and I will send to you detailed information on how this business would be successfully transacted. Be informed that nobody knows about the secret of this fund except us, and we know how to carry out the entire transaction. So I am compelled to ask, that you will stand on our behalf and receive this fund into any account that is solely controlled by you.

We will compensate you with 30% of the total amount involved as gratification for being our partner in this transaction. Reply to my private email as stated: honreemebrahimal-hashimi@yandex.com

Regards,
Ms. Reem Ebrahim Al-Hashimi.

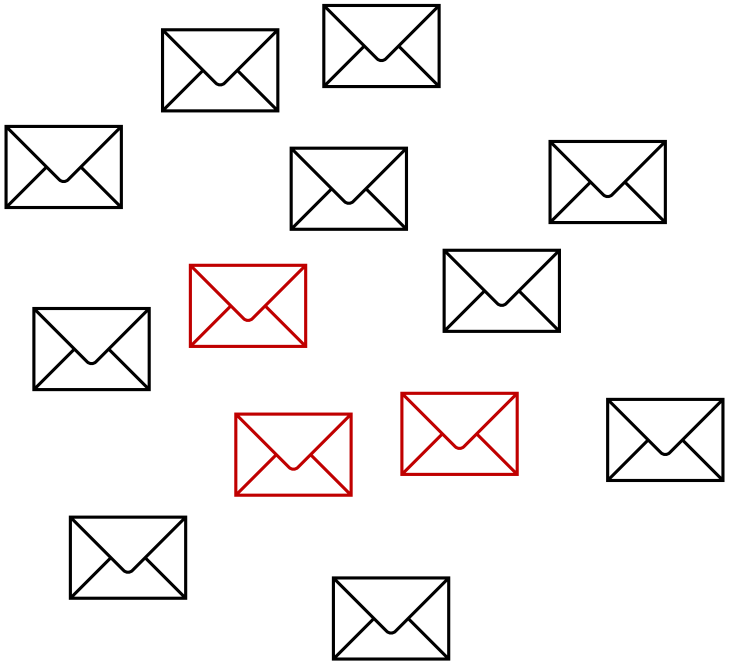
We classify the spams based on our **experience**

Hello Dear,
Am a dying woman here in the hospital, i was diagnose as a cancer patient 2 years ago. I am A business woman
dealing with Gold Exportation. I have a charitable and unfulfillment project that am about to handover to you, if you are interested please Reply
hope to hear from you.
Regard
Mrs.Elizabeth Kari,

We classify the spams based on our **experience**

Google Customer Reward Program
Congratulations!
1 October 2019
We would like to thank you for using our products for so long!
Every day, we select a small group of Google users and give them the chance to get valuable gifts from our partners and sponsors. In this way, we thank you for choosing to use Google.
You can win a new Samsung Galaxy S10, iPhone X or iPad Air 2. All you have to do to get a gift is to answer the following 7 questions.
This survey allows us to better understand the users and improve our products. It will not take more than 30 seconds of your time.
Note: 100 randomly selected users have received this invitation and the number of gifts is limited.
You have 1 minutes and 22 seconds to answer the following questions before we give the gift to another lucky user! I wish you success!

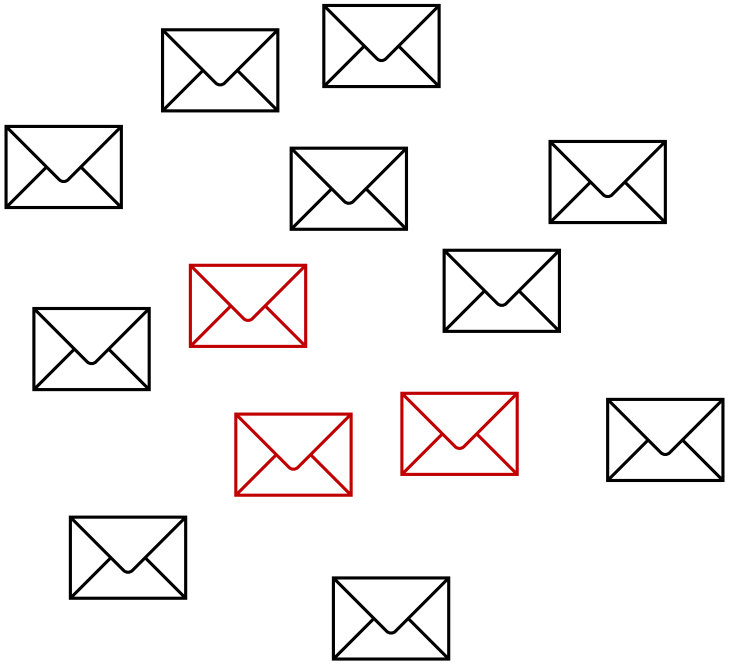
Probabilistic thinking

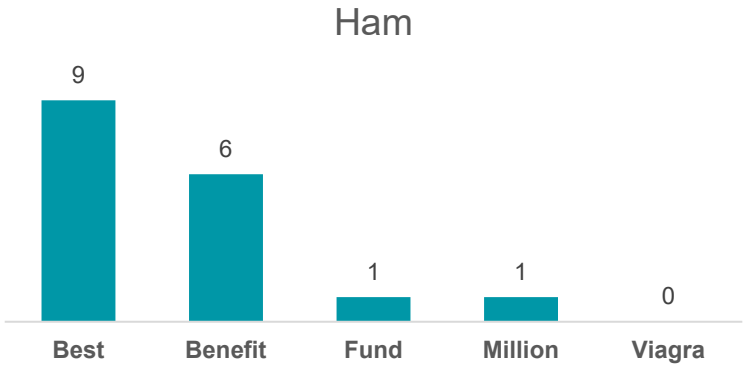


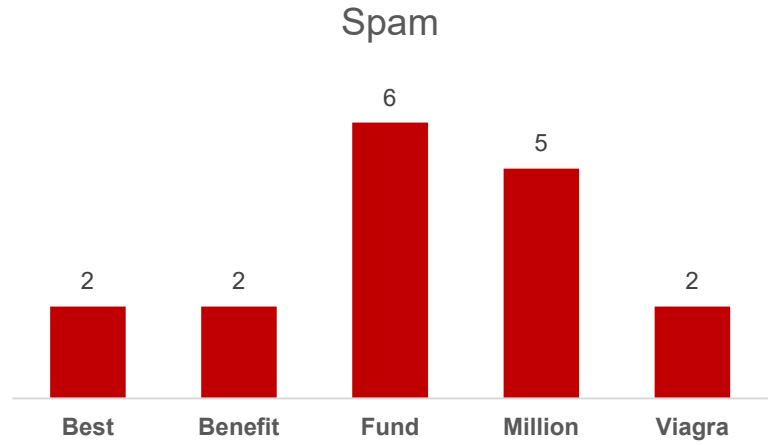
$$P(\text{Ham}) = \frac{10}{13}$$

$$P(\text{Spam}) = \frac{3}{13}$$

Word frequency







Word frequency

Hello Dear,

I'd like to offer the best chance to buy Viagra.
If you are interested …

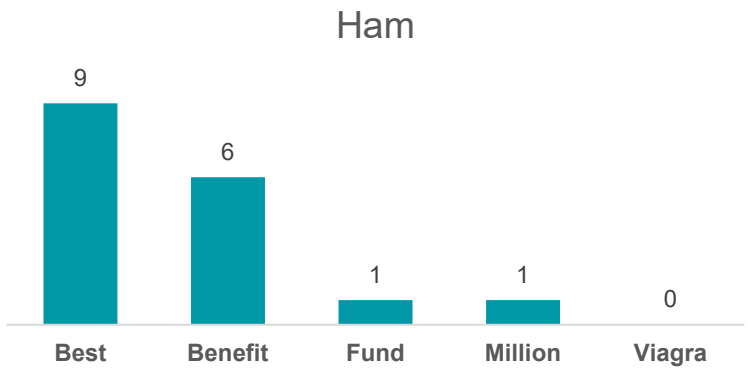
The benefit you get is that …

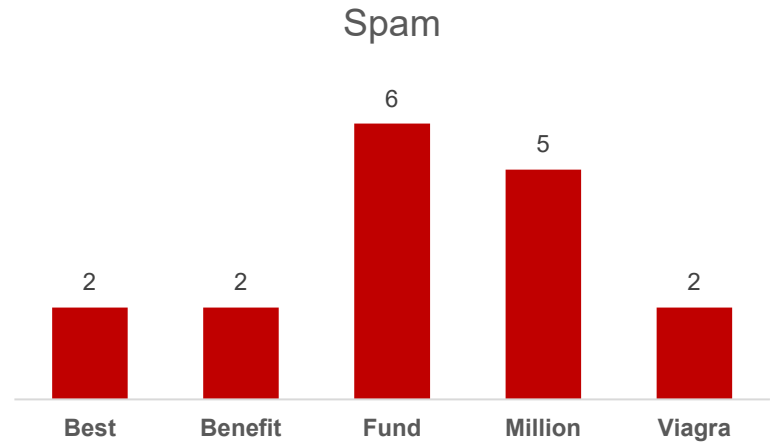
Also want to suggest a good fund..

Amount of 100 Million US$ …

hope to hear from you.

Yours sincerely,
….





Word frequency

Hello Dear,

I'd like to offer the best chance to buy Viagra.
If you are interested …

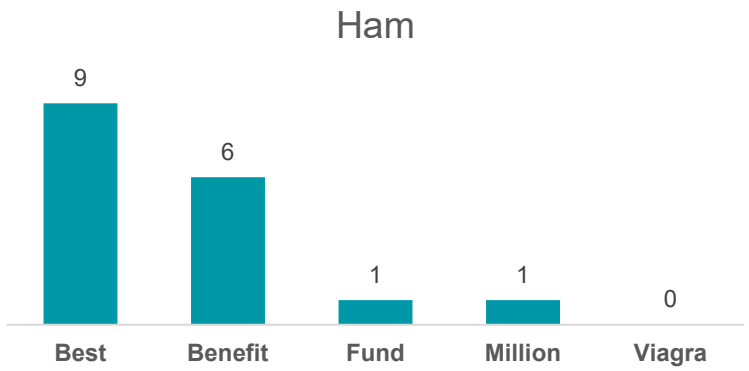
The benefit you get is that …

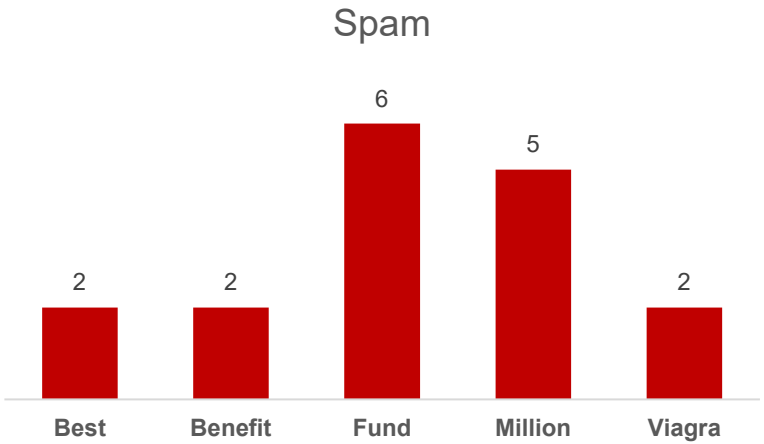
Also want to suggest a good fund..

Amount of 100 Million US$ …

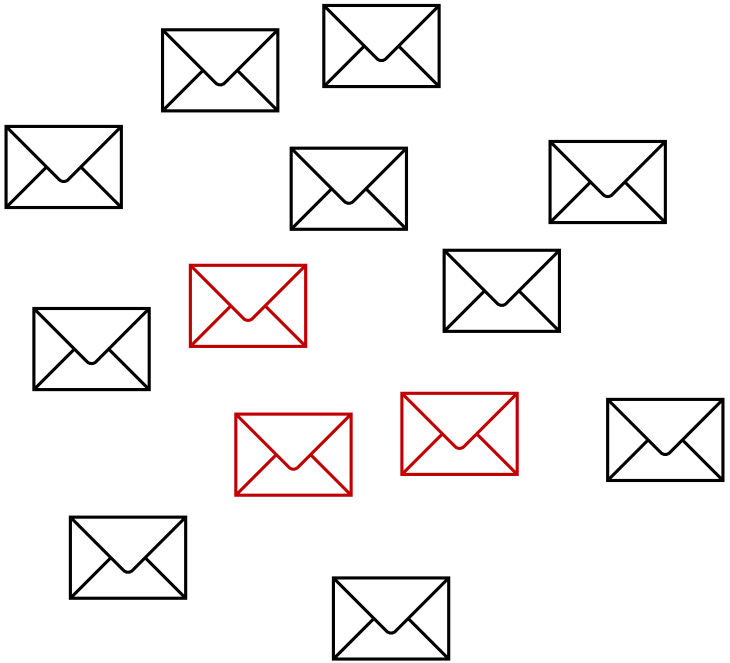
hope to hear from you.

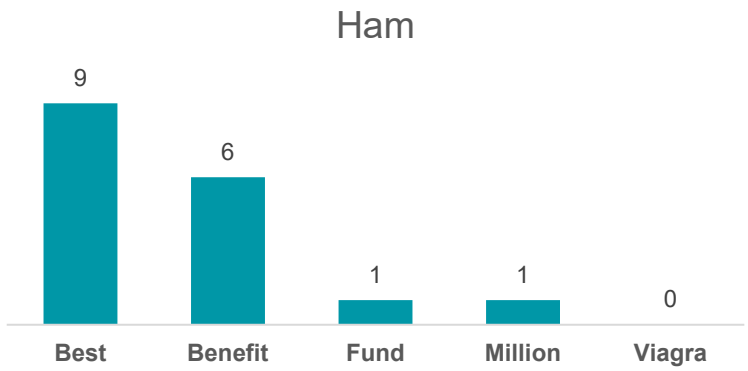
Yours sincerely,
….





Word frequency





$$P(Best \mid Ham) = \frac{9}{17}$$

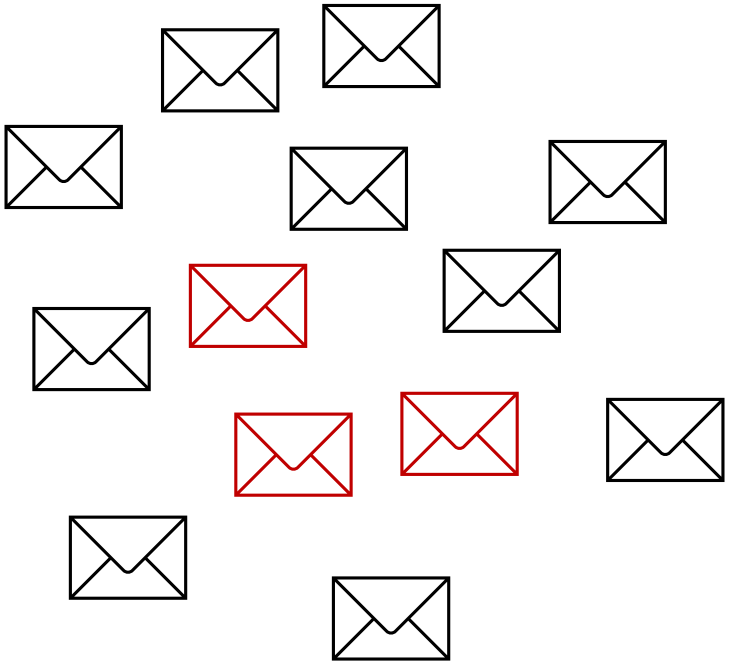
$$P(Benefit \mid Ham) = \frac{6}{17}$$

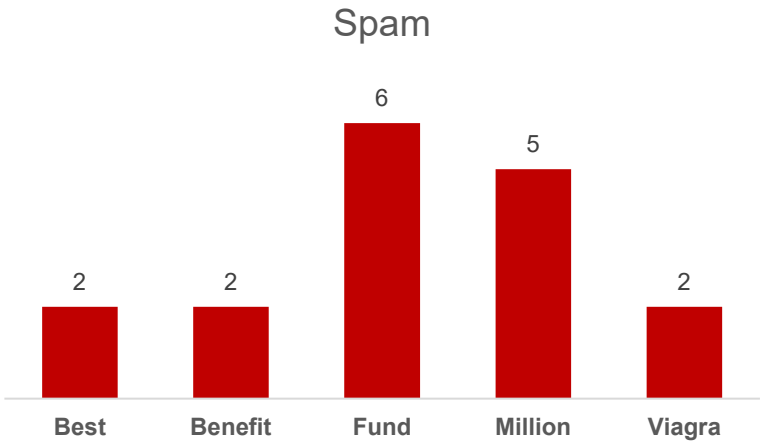
$$P(Fund \mid Ham) = \frac{1}{17}$$

$$P(Million \mid Ham) = \frac{1}{17}$$

$$P(Viagra \mid Ham) = \frac{0}{17}$$

Word frequency





$$P(Best \mid Spam) = \frac{2}{17}$$

$$P(Benefit \mid Spam) = \frac{2}{17}$$

$$P(Fund \mid Spam) = \frac{6}{17}$$

$$P(Million \mid Spam) = \frac{5}{17}$$

$$P(Viagra \mid Spam) = \frac{2}{17}$$

Bayes Theorem

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

$$P(A \cap B) = P(B \mid A) \times P(A)$$

Likelihood

Prior knowledge

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

Marginal likelihood

Bayes Theorem

Likelihood

Prior knowledge

$$P(Spam \mid fund) = \frac{P(fund|Spam) \times P(Spam)}{P(fund)}$$

Marginal likelihood

Bayes Theorem

Likelihood

Prior knowledge

$$P(Ham \mid fund) = \frac{P(fund|Ham) \times P(Ham)}{P(fund)}$$

Marginal likelihood

Bayes Theorem

$$P(Ham \mid fund) \quad \textbf{vs.} \quad P(Spam \mid fund)$$

$$\frac{P(fund|Ham) \times P(Ham)}{P(fund)} \qquad \frac{P(fund|Spam) \times P(Spam)}{P(fund)}$$

$$P(fund|Ham) = \frac{1}{17} \qquad P(fund|Spam) = \frac{6}{17}$$

$$P(Ham) = \frac{10}{13} \qquad P(Spam) = \frac{3}{13}$$

Naïve Bayes Classifier: how it works?

Hello Dear,

…

The benefit you get is that …

Also want to suggest a good fund..

Amount of 100 Million US$ … fund …

hope to hear from you.

Yours sincerely,

….

$$P(Ham \mid Evidence) \quad \text{VS.} \quad P(Spam \mid Evidence)$$

$$P(benefit \mid Ham) \times P(fund \mid Ham)^2 \times P(Milliion \mid Ham) \times P(Ham)$$

$$P(Benefit \mid Ham) = \frac{6}{17}$$

$$P(Fund \mid Ham) = \frac{1}{17}$$

$$P(Million \mid Ham) = \frac{1}{17}$$

$$P(Ham) = \frac{10}{13}$$

Naïve Bayes Classifier: how it works?

Hello Dear,

…

The benefit you get is that …

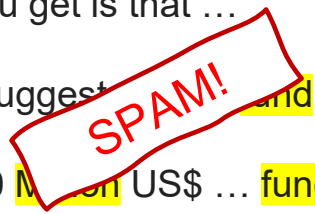
Also want to suggest … fund..

Amount of 100 M…on US$ … fund …

hope to hear from you.

Yours sincerely,

….



$$P(Ham \mid Evidence) \quad \text{Vs.} \quad P(Spam \mid Evidence)$$

$$P(benefit \mid Spam) \times P(fund \mid Spam)^2 \times P(Milliion \mid Spam) \times P(Spam)$$

$$P(Benefit \mid Spam) = \frac{2}{17}$$

$$P(Fund \mid Spam) = \frac{6}{17}$$

$$P(Million \mid Spam) = \frac{5}{17}$$

$$P(Spam) = \frac{3}{13}$$

Naïve Bayes Classifier

In Titanic cruise, how can we classify whether survived or not a female 1st class was?

$P(Survived \mid Femle \ \& \ 1st \ Class)$

$P(Femle|Survived) \times P(1st \ Class|Survived) \times P(Survived)$

$$\frac{240}{300} \times \frac{120}{300} \times \frac{300}{600} = \frac{48}{300}$$

$P(Not \ survived \mid Femle \ \& \ 1st \ Class)$

$P(Femle|Not \ survived) \times P(1st \ Class|Not \ survived) \times P(Not \ survived)$

$$\frac{60}{300} \times \frac{80}{300} \times \frac{300}{600} = \frac{8}{300}$$

	Female	Male
1st	(90/10)	(30/70)
2nd	(80/20)	(20/80)
3rd	(70/30)	(10/90)

Total # passengers: 600
Survived: 300
Not survived: 300
Female & survived: 240
Female & not survived: 60
1st Class & survived: 120
1st Class & not survived: 80

$P(Survived)$ = 300 / 600 = 0.5