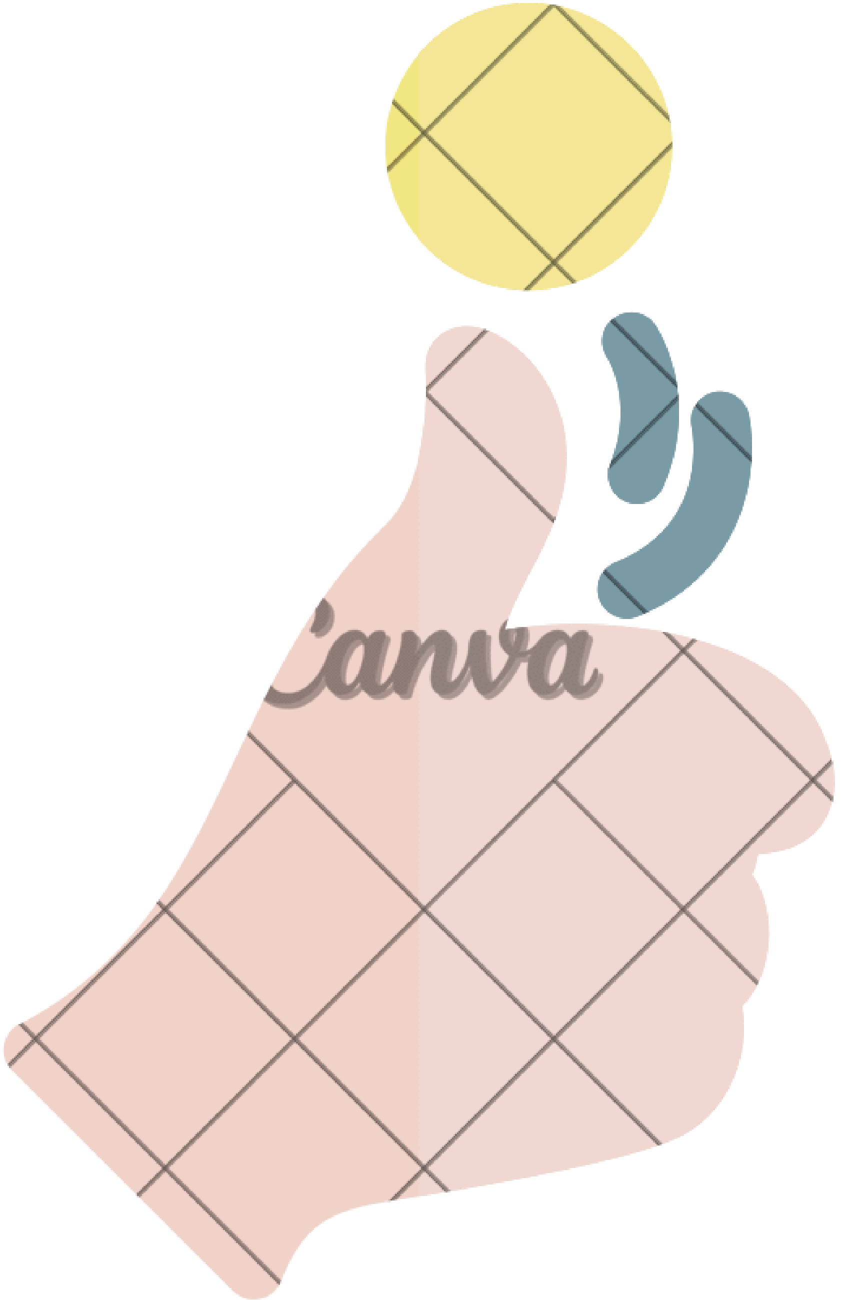
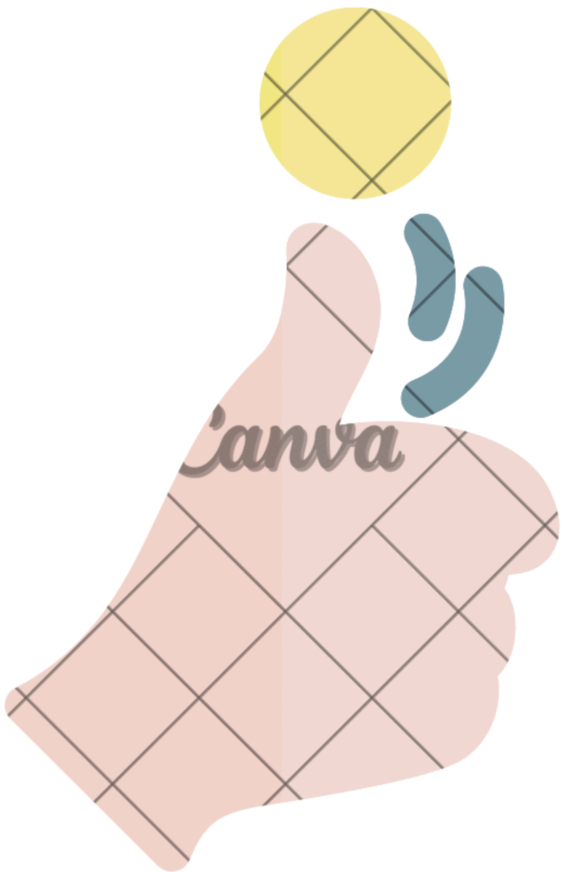
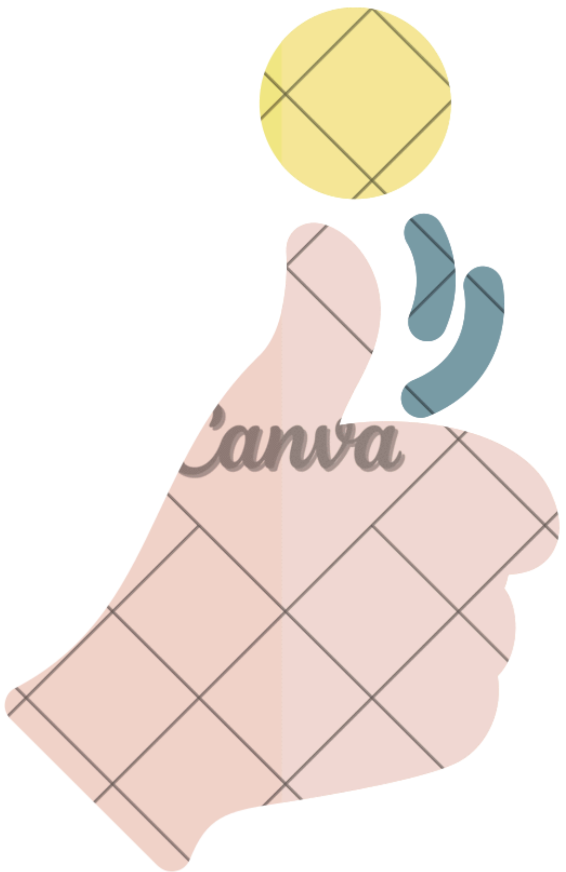
Naive Bayes

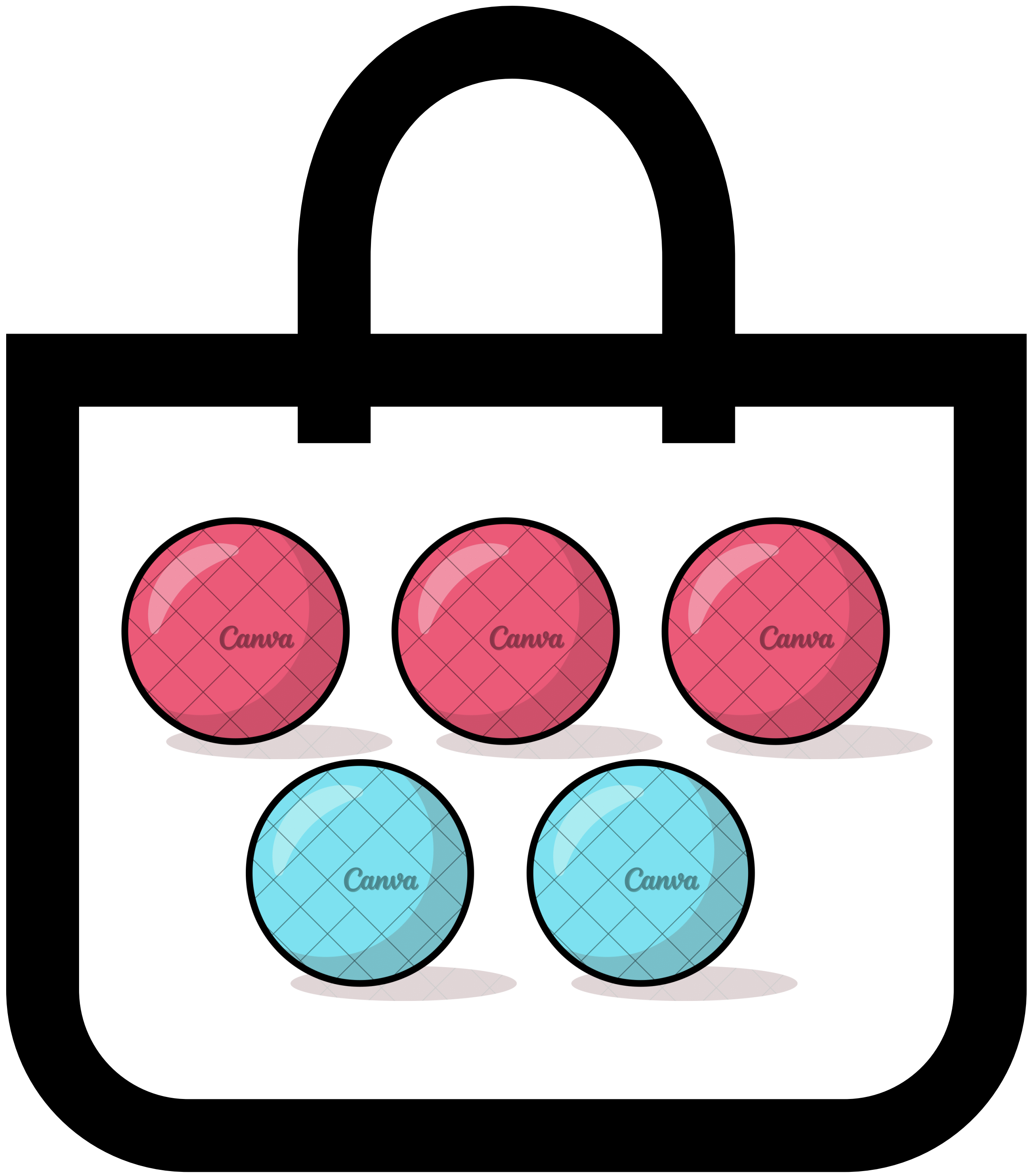
Naive Bayes

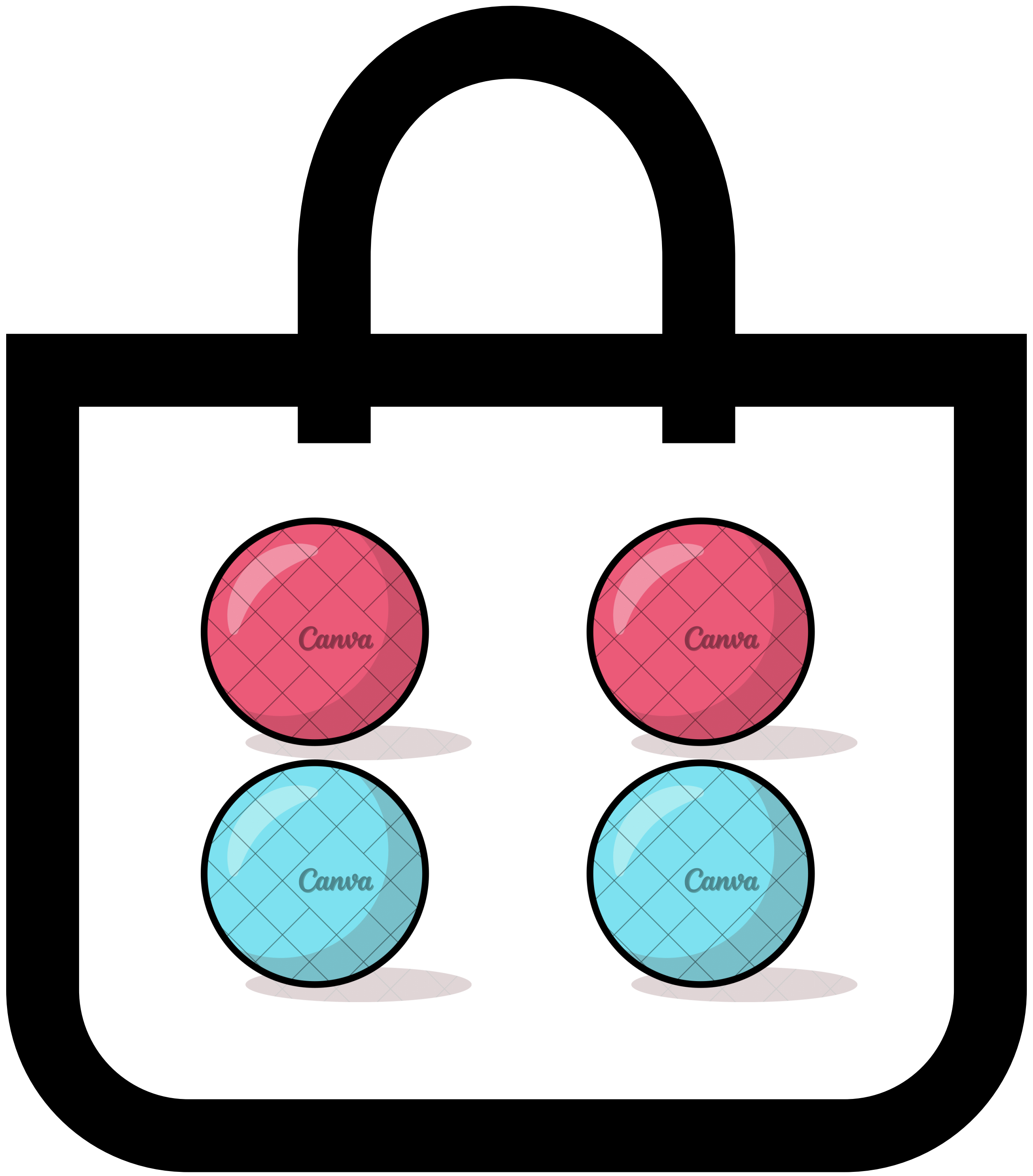



Canva


Canva

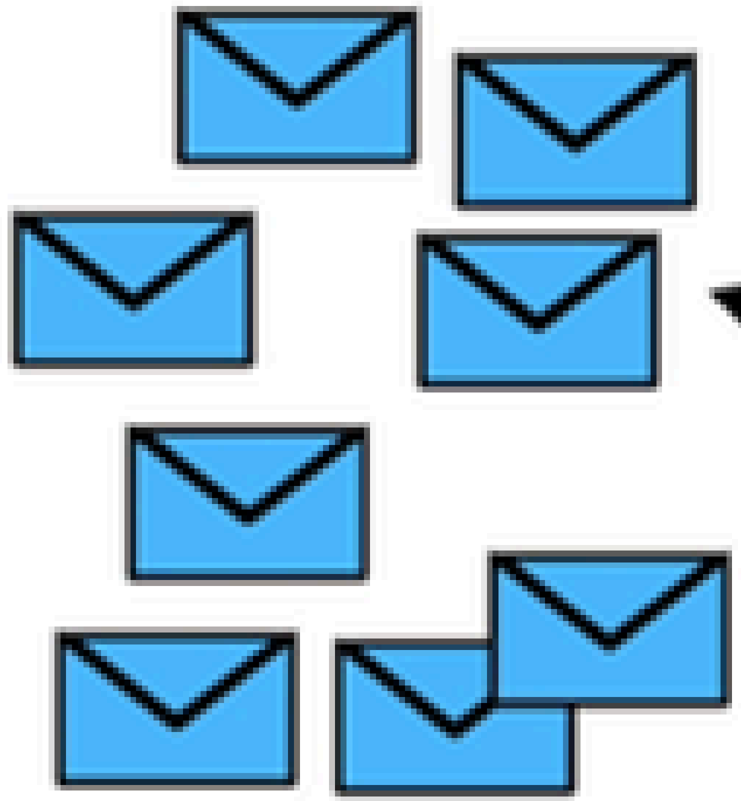
Conditional probability



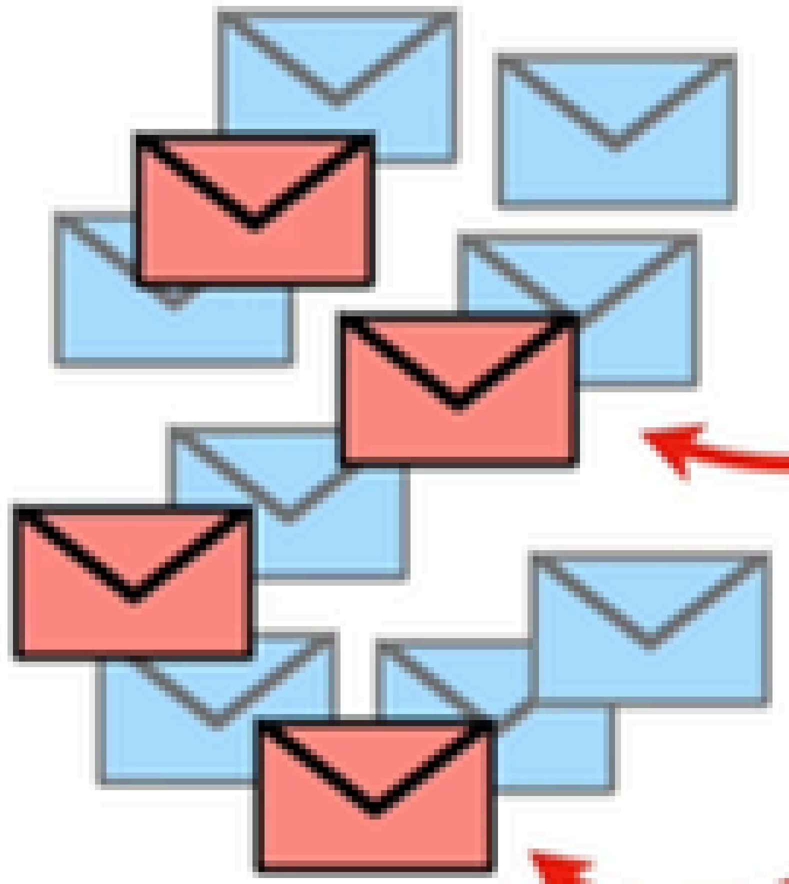


Thomas Bayes

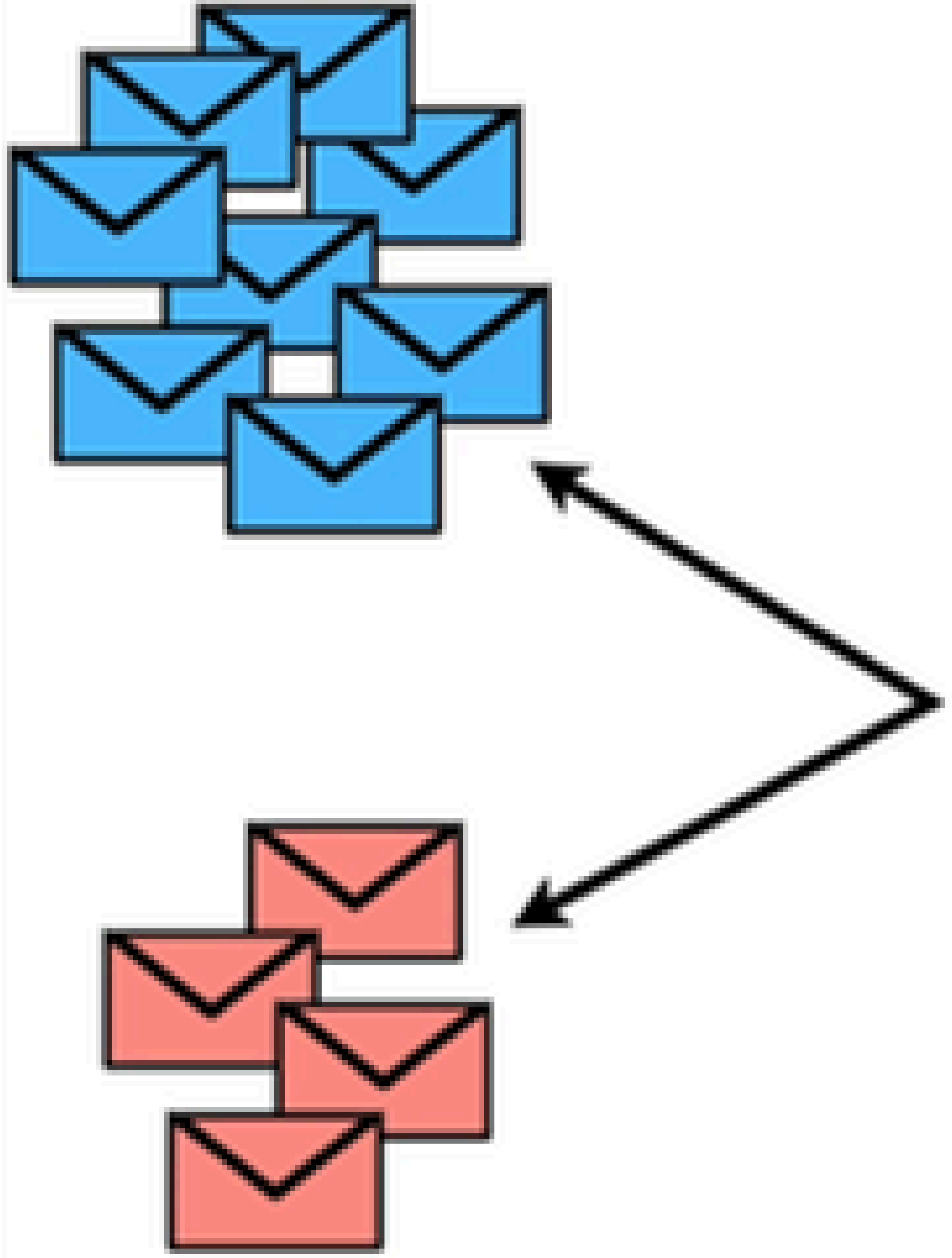
$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$



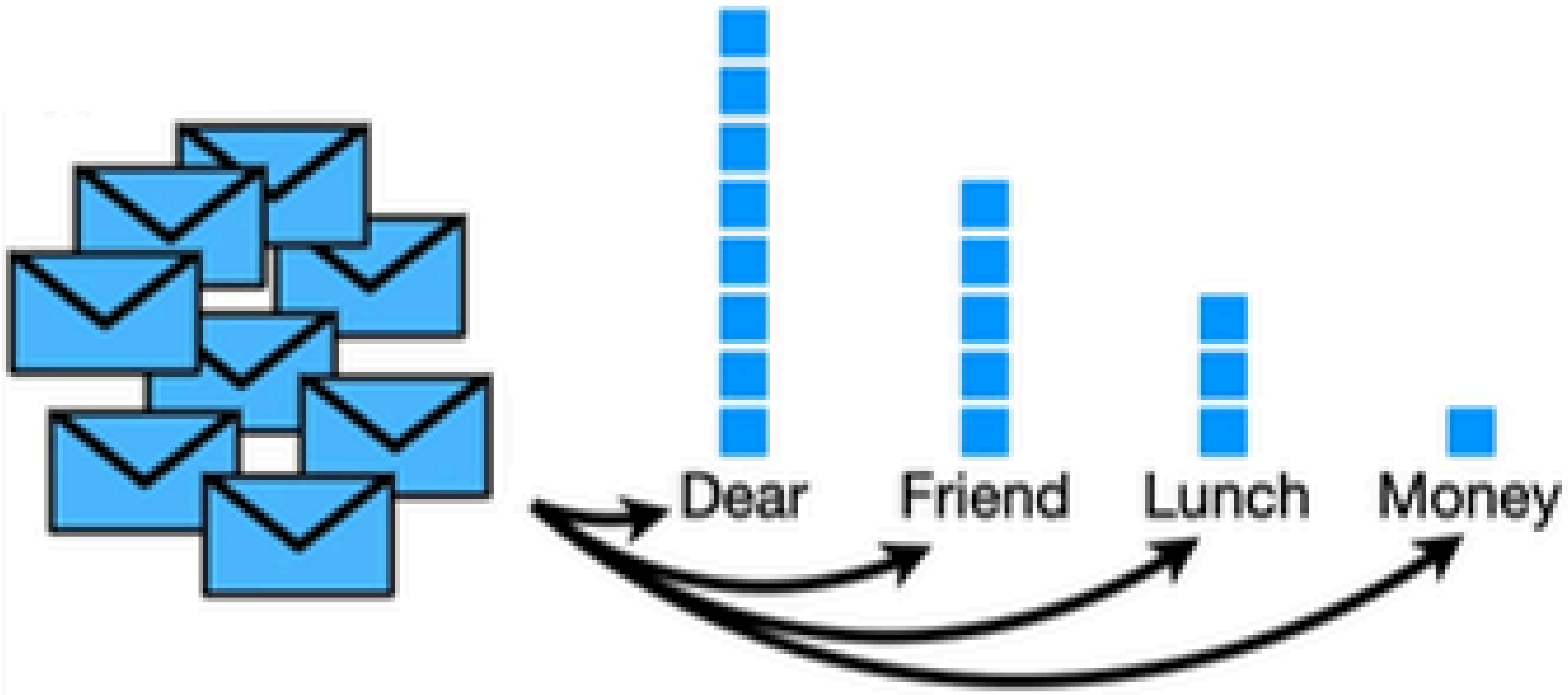
Imagine we received **normal messages** from friends and family…



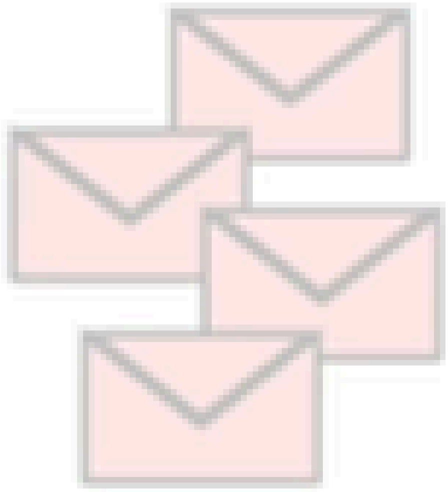
…and we also received **spam** (unwanted messages that are usually scams or unsolicited advertisements)…

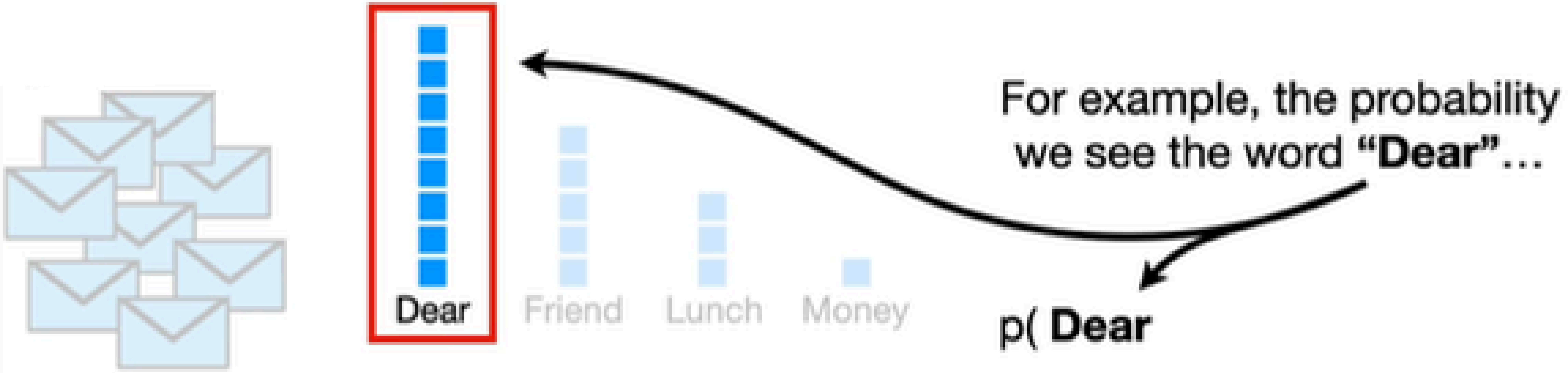


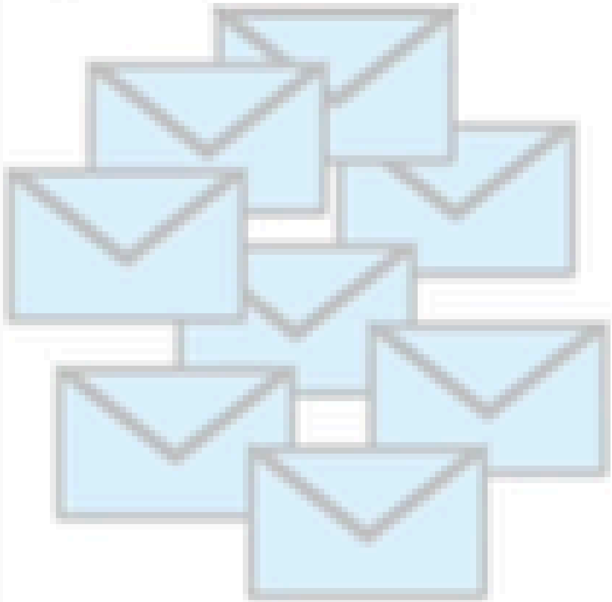
…and we wanted to filter out the **spam** messages.

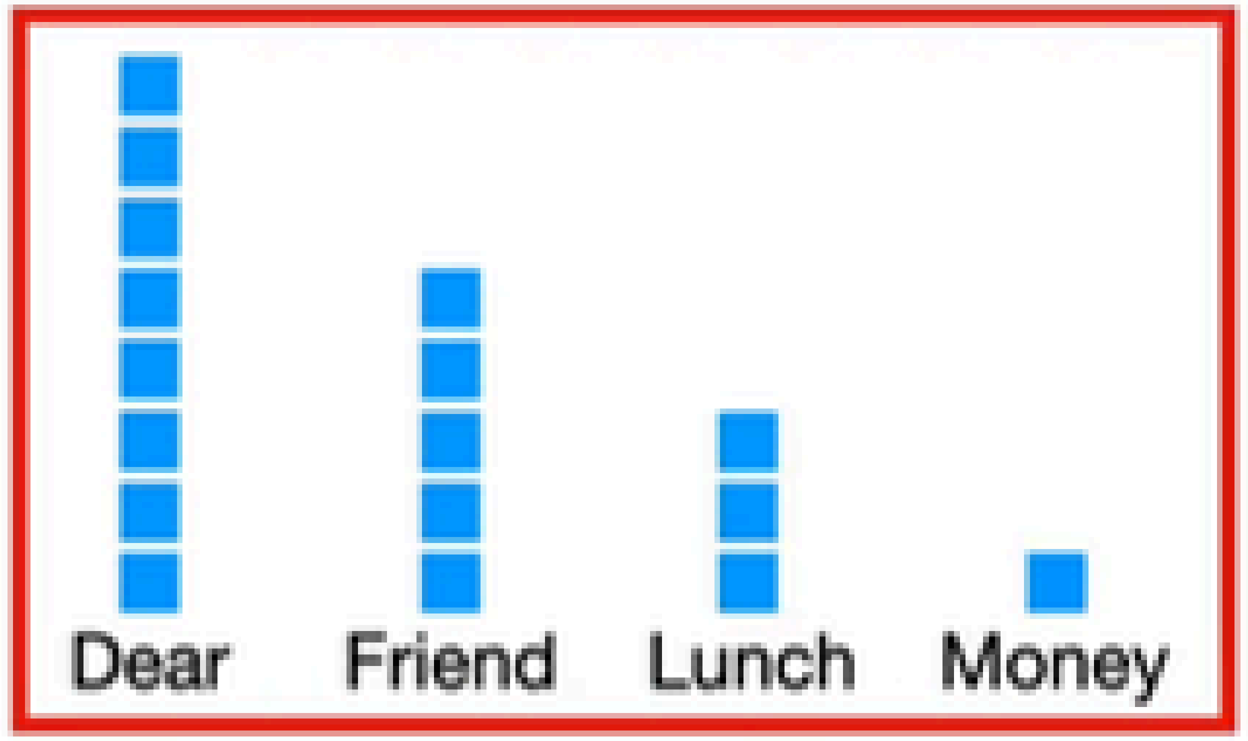


So, the first thing we do is make a **histogram** of all the words that occur in the normal messages from friends and family.



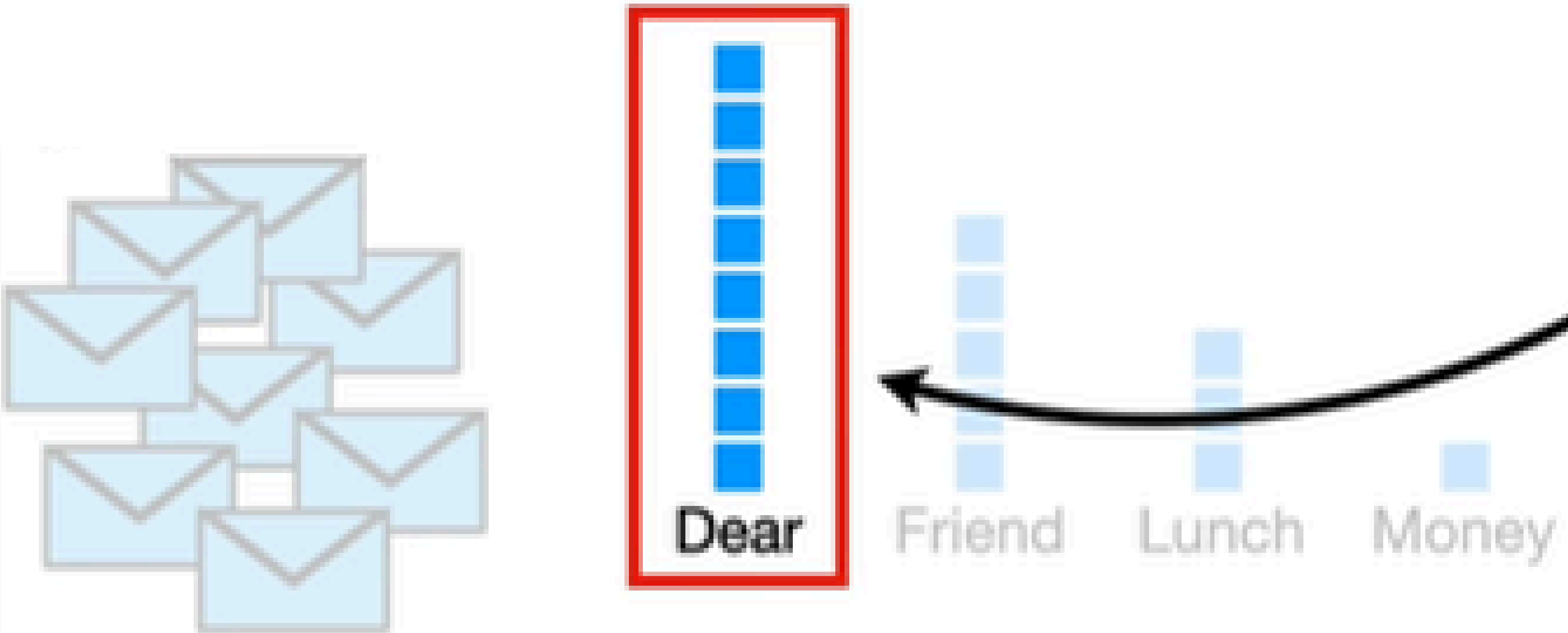

Dear
Friend
Lunch
Money
For example, the probability we see the word **"Dear"**…
p(**Dear**





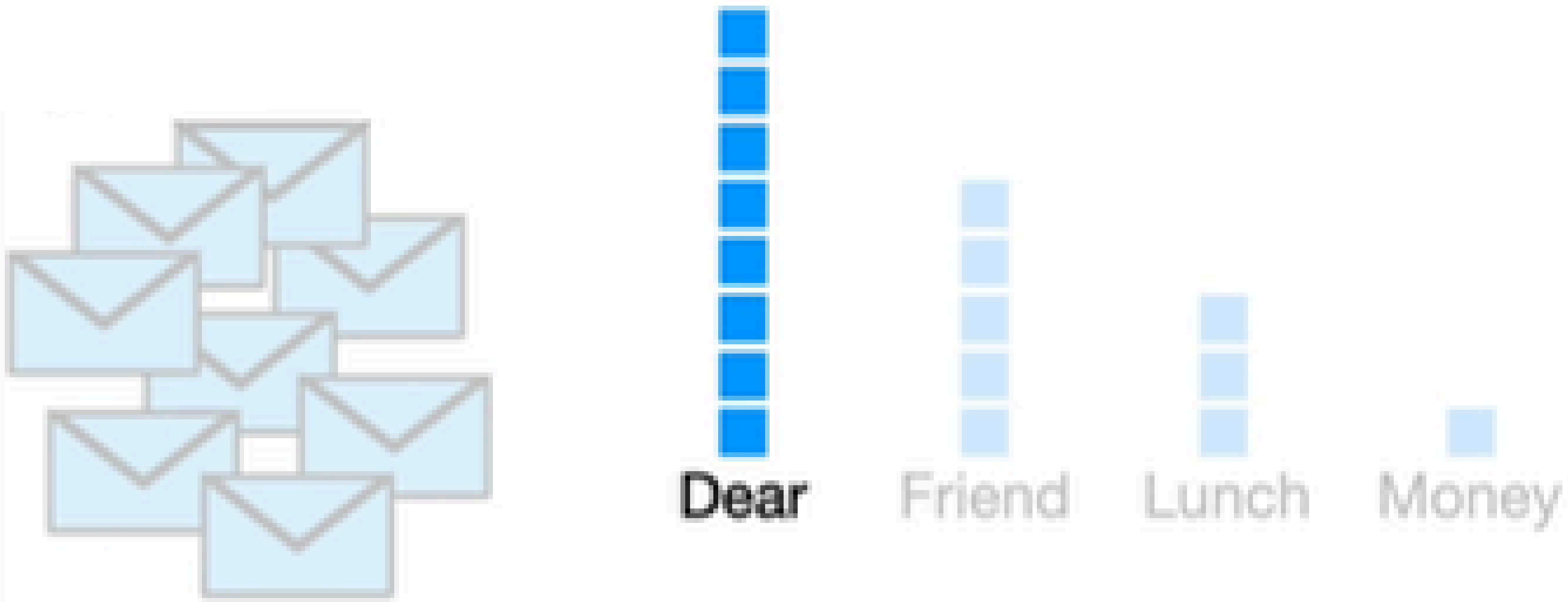
…given that we saw it in a **normal message**…

p(**Dear** | **Normal**)



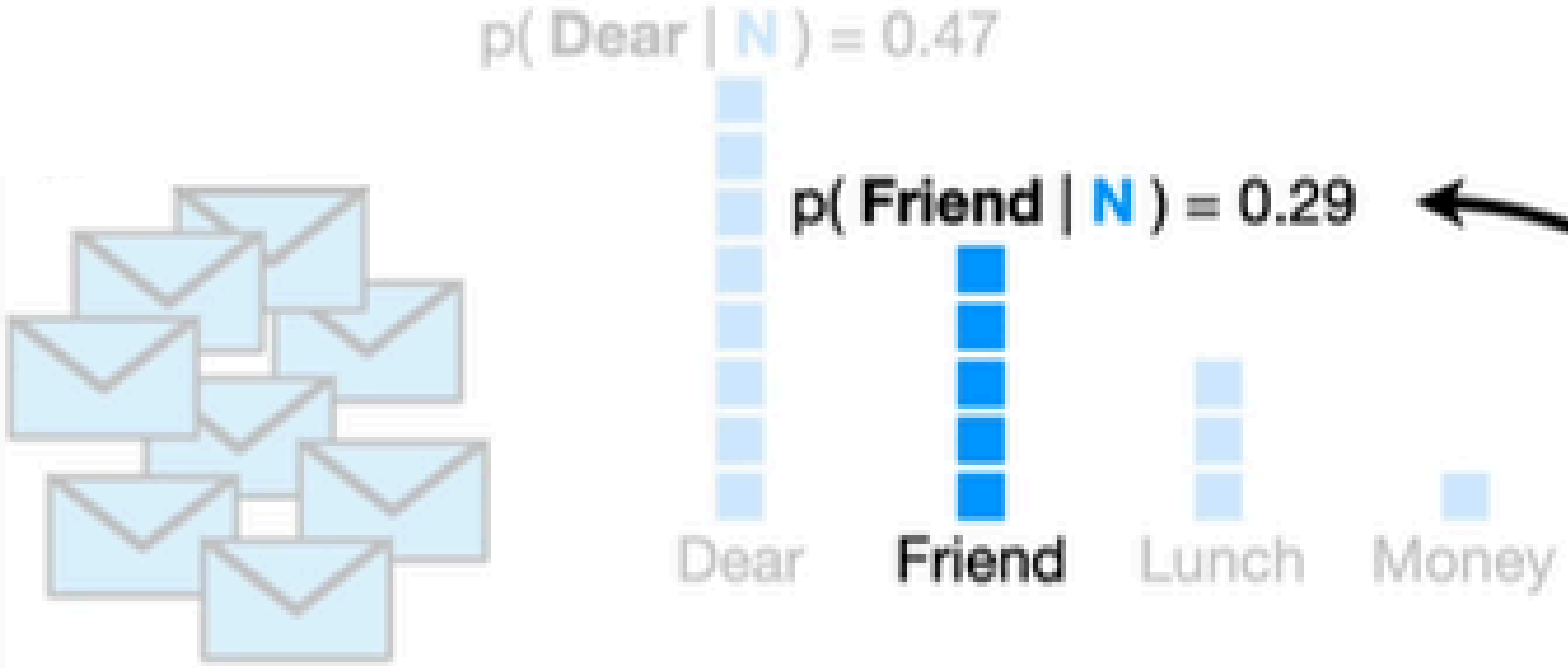
…is **8**, the total number of times **Dear** occurred in **normal messages**…

$$p(\,\mathbf{Dear} \mid \mathbf{Normal}\,) = \frac{8}{17}$$



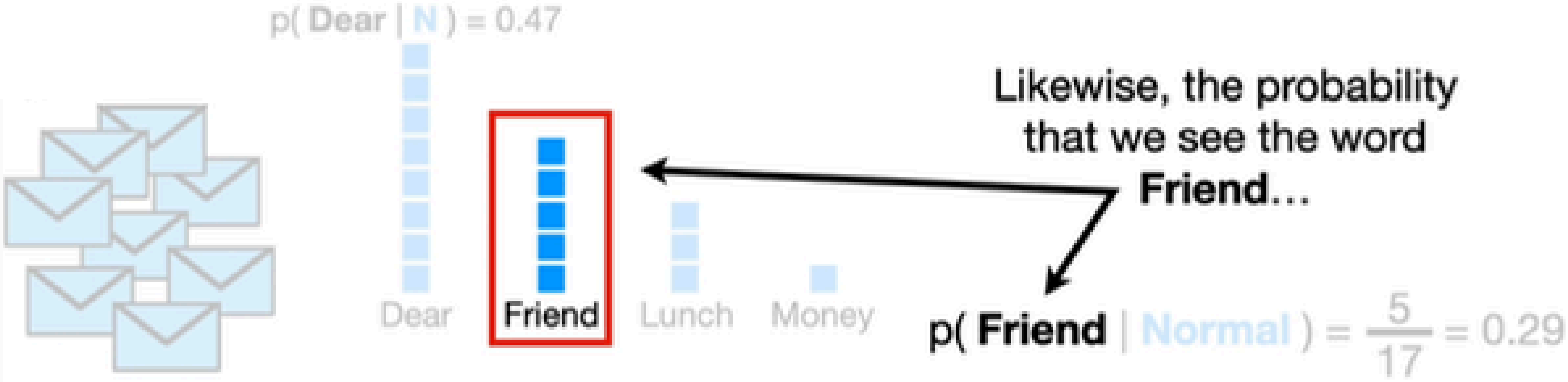
And that gives us **0.47**.

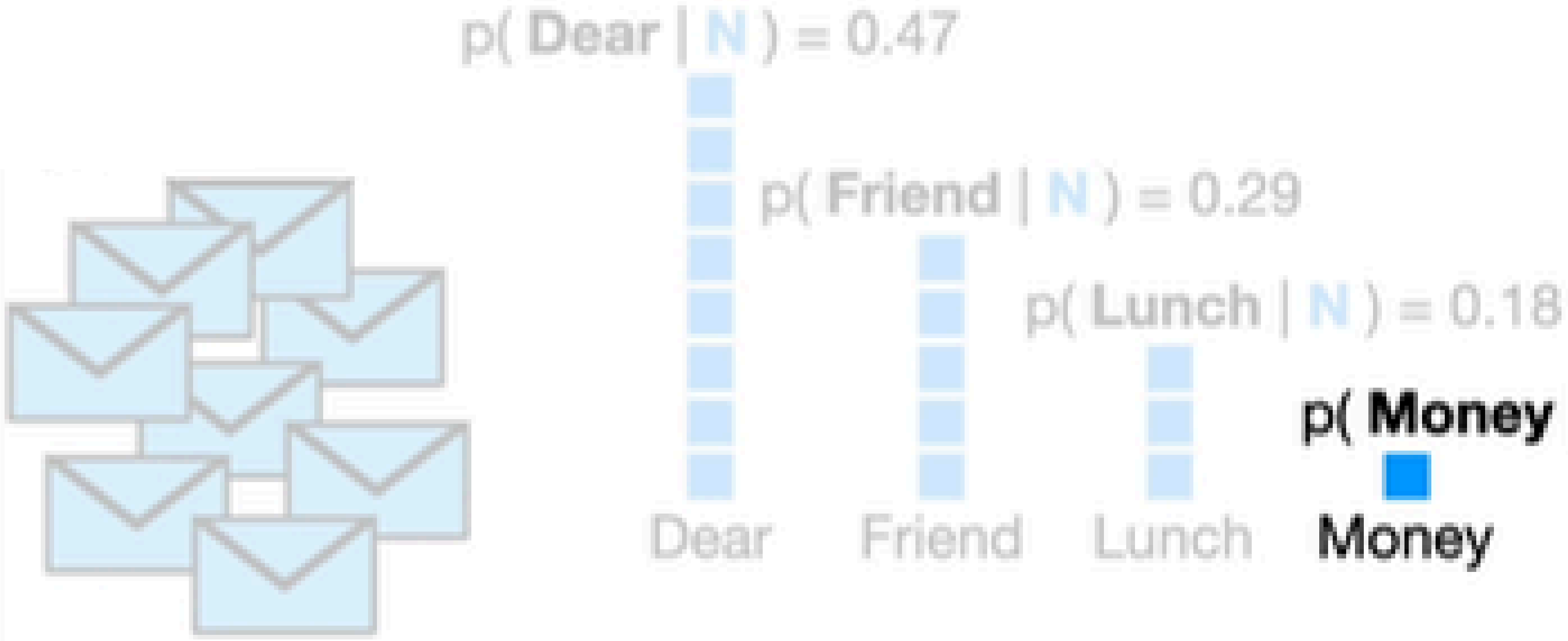
$$p(\textbf{Dear} \mid \textbf{Normal}) = \frac{8}{17} \boxed{= 0.47}$$



So let's put that over the word **Friend**, so we don't forget it.

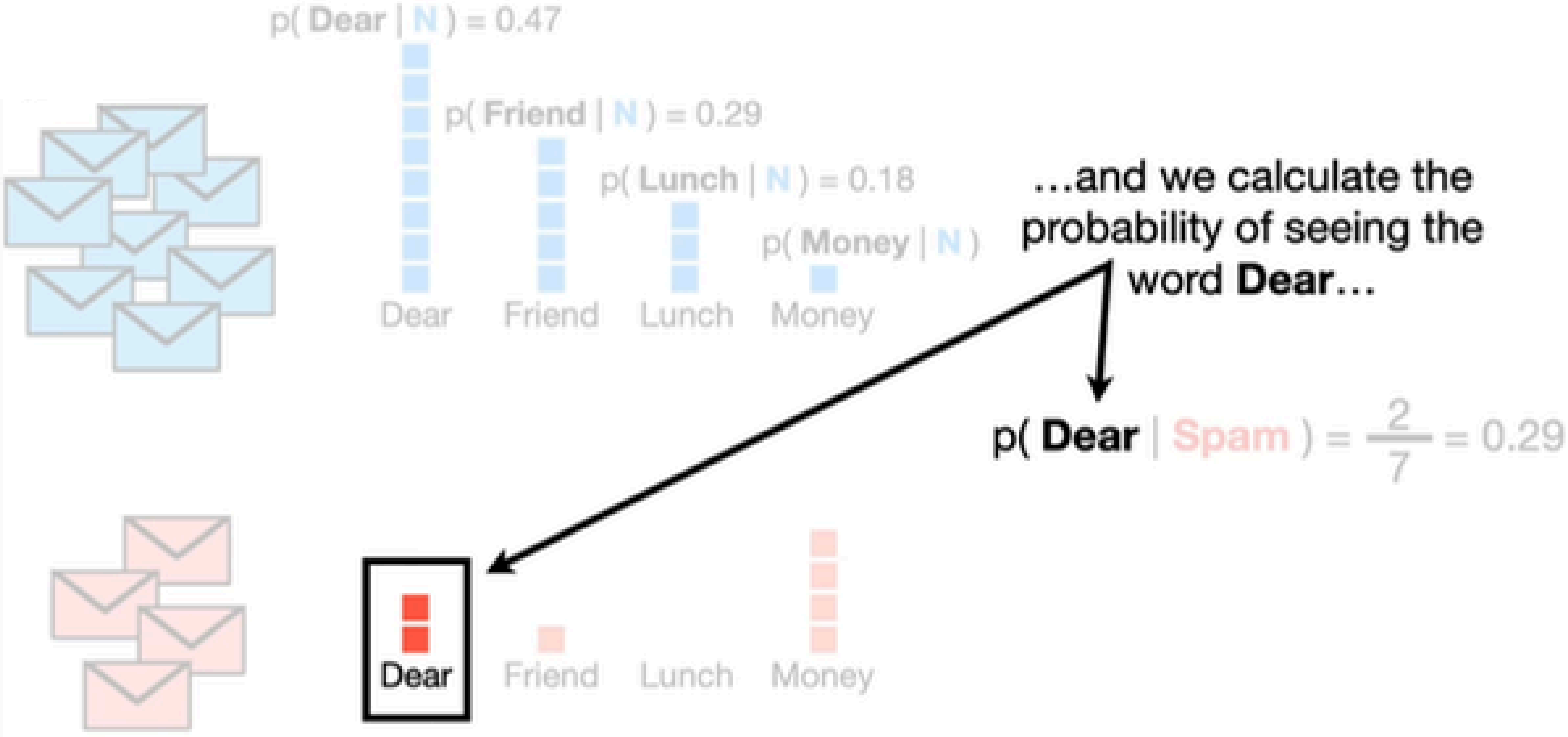
$$p(\textbf{ Friend } | \textbf{ Normal }) = \frac{5}{17} = 0.29$$


p(Dear | N) = 0.47
Dear
Friend
Lunch
Money
Likewise, the probability that we see the word Friend...
p(Friend | Normal) = 5/17 = 0.29

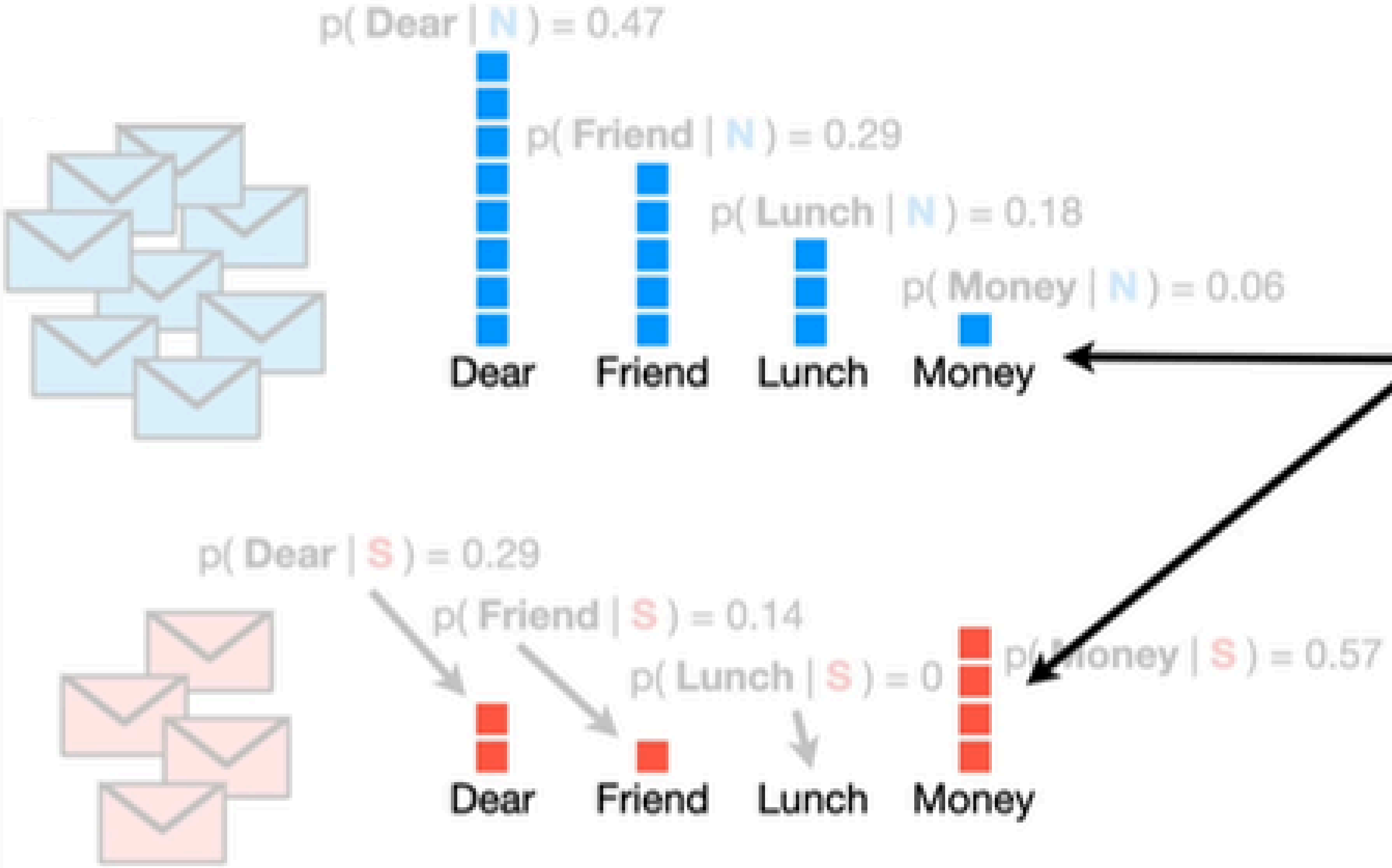


…and the probability that we see the word **Money**, given that it is in a **normal message** is **0.06**.

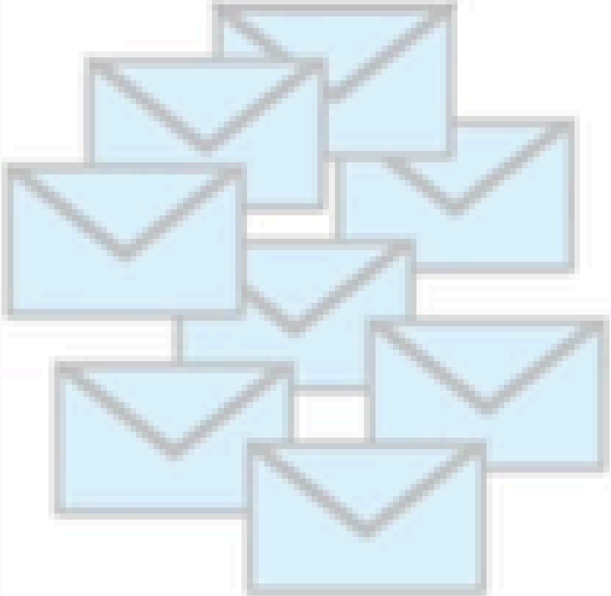
$$p(\textbf{Money} \mid \textbf{Normal}) = \frac{1}{17} = 0.06$$



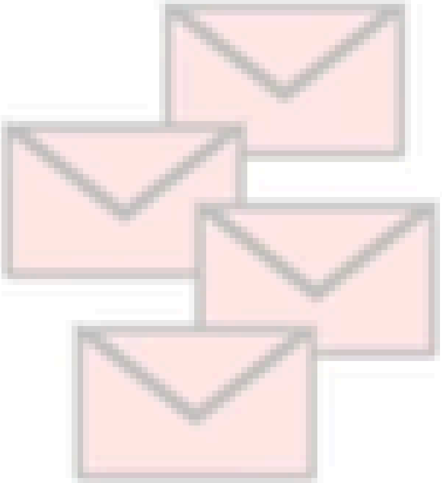
p(Dear | N) = 0.47
p(Friend | N) = 0.29
p(Lunch | N) = 0.18
p(Money | N)
Dear
Friend
Lunch
Money
…and we calculate the probability of seeing the word Dear…
p(Dear | Spam) = 2/7 = 0.29
Dear
Friend
Lunch
Money



Now, because these histograms are taking up a lot of space, let's get rid of them, but keep the probabilities.



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

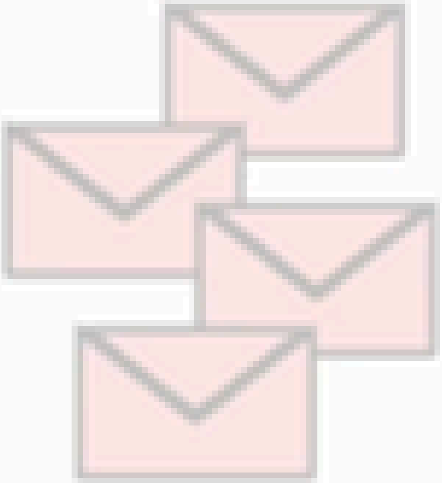
Because we have calculated the probabilities of discrete, individual words, and not the probability of something continuous, like weight or height, these **Probabilities** are also called **Likelihoods**.

Dear Friend

We start with an initial guess about the probability that any message, regardless of what it says, is a **normal message**.

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

p(**N**)



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

Dear Friend

The guess can be any probability that we want, but a common guess is estimated from the training data.

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

p(**N**)

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

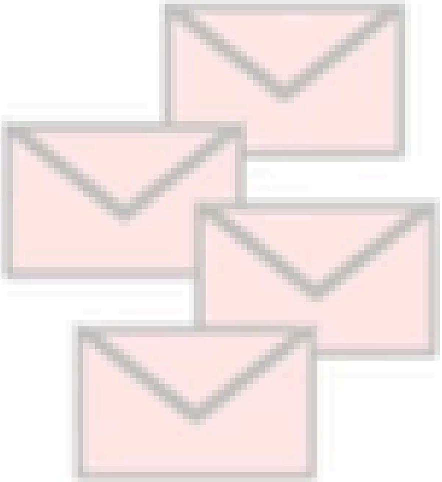
Dear Friend

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

p(**N**) = 0.67

So let's put that under the **normal messages** so we don't forget it.

$$p(\mathbf{N}) = \frac{8}{8 + 4} = 0.67$$



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

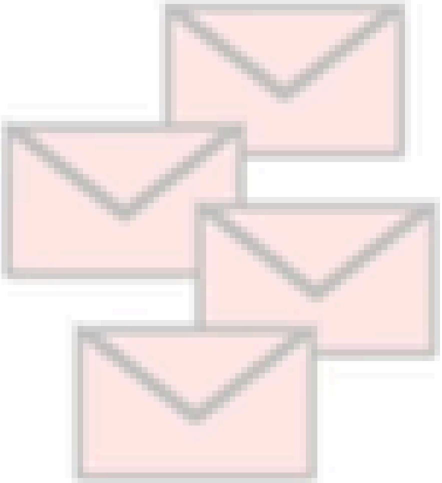
Dear Friend

Now we multiply that initial guess by the probability that the word **Dear** occurs in a **normal message**…

p(N) = 0.67

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

p(N) × p(**Dear** | N)



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

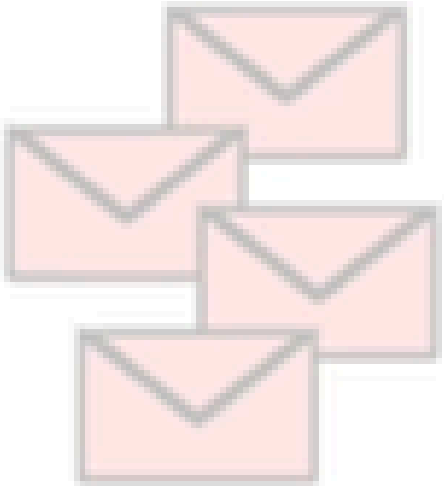
Dear Friend

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

p(N) = 0.67

Now we just plug in the values that we worked out earlier and do the math...

p(N) × p(**Dear** | N) × p(**Friend** | N)



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

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

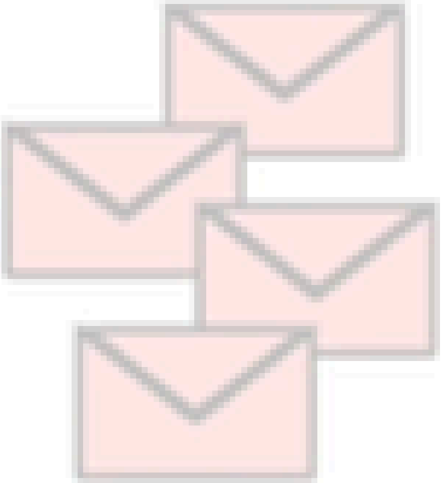
Dear Friend

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

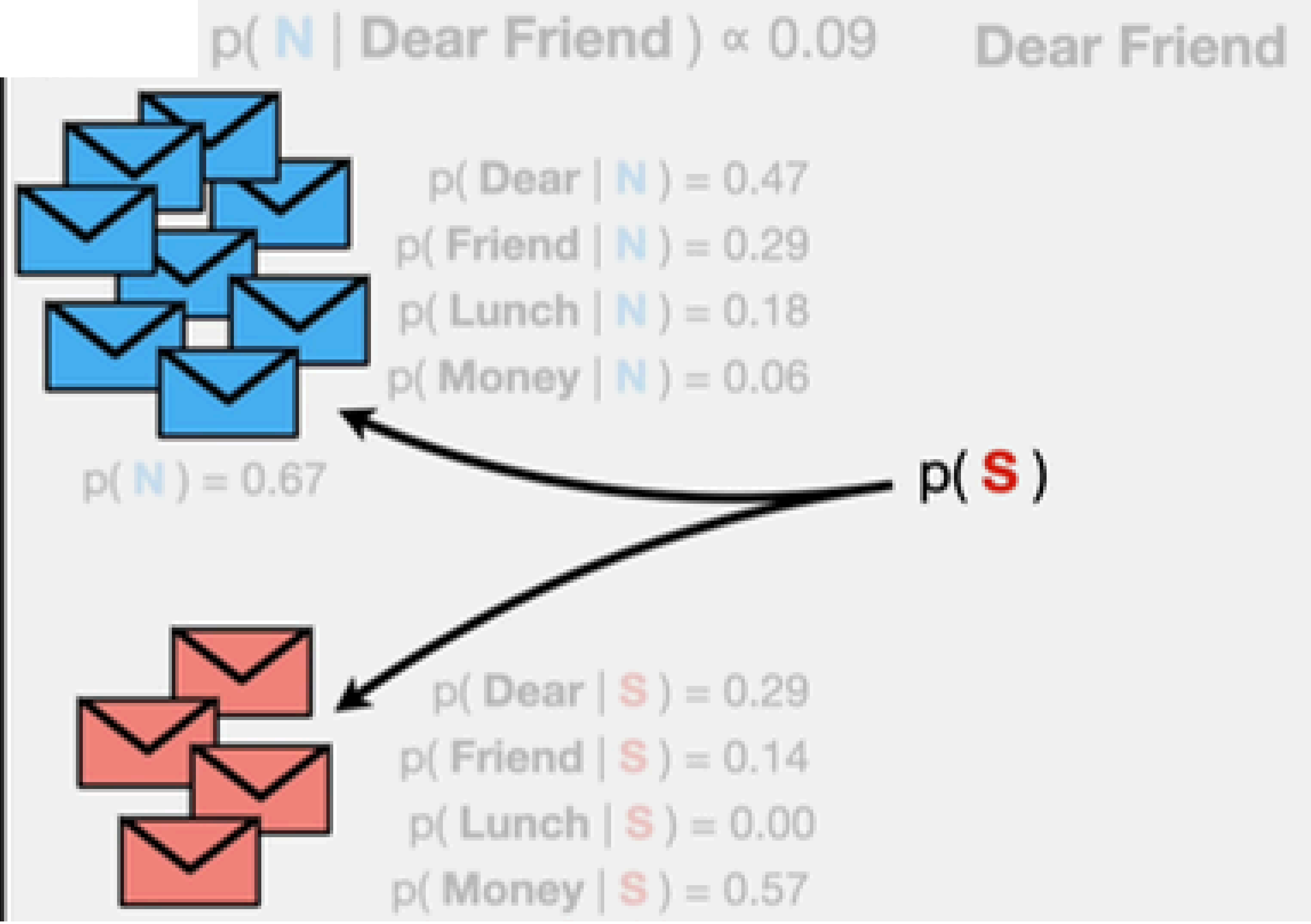
$p(\text{N}) = 0.67$

So let's put that on top of the **normal messages** so we don't forget.

$0.67 \times 0.47 \times 0.29 = 0.09 \propto p(\text{N} \mid \text{Dear Friend})$



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

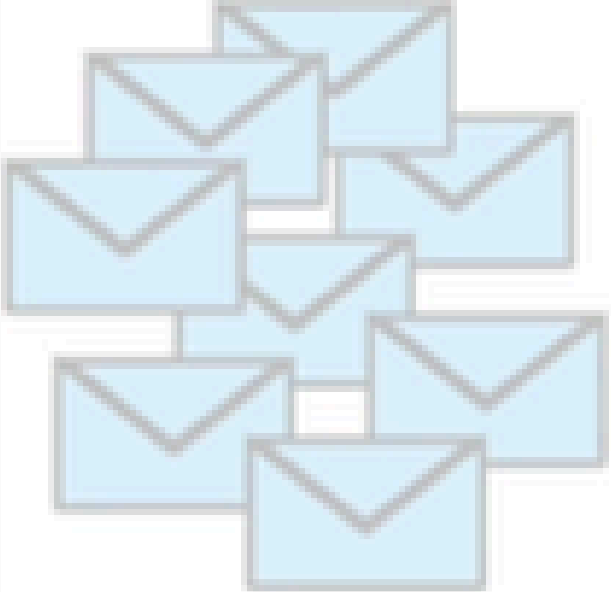


And just like before, the guess can be any probability that we want, but a common guess is estimated from the training data.

p(N | Dear Friend) ∝ 0.09

Dear Friend

So let's put that under the spam so we don't forget it.



p(Dear | N) = 0.47

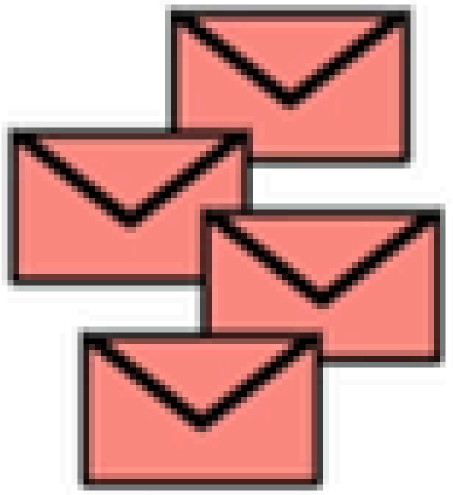
p(Friend | N) = 0.29

p(Lunch | N) = 0.18

p(Money | N) = 0.06

p(N) = 0.67

$$p(S) = \frac{4}{4 + 8} = 0.33$$



p(Dear | S) = 0.29

p(Friend | S) = 0.14

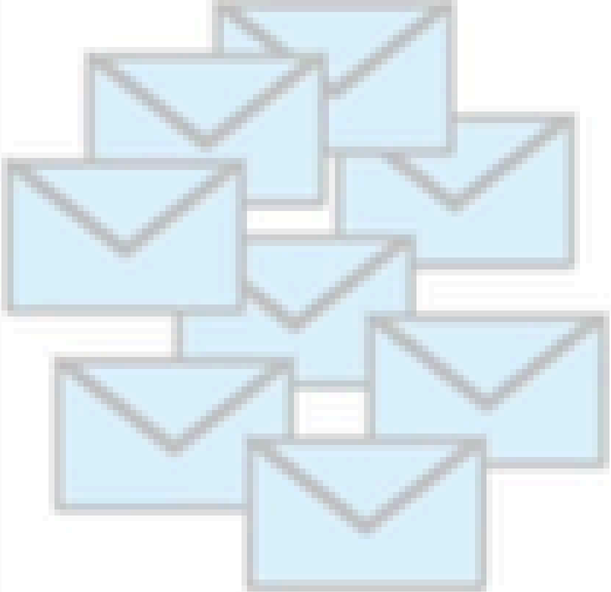
p(Lunch | S) = 0.00

p(Money | S) = 0.57

p(S) = 0.33

p(N | Dear Friend) ∝ 0.09

Dear **Friend**

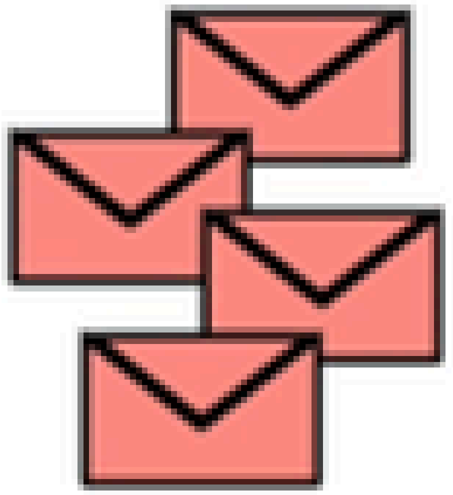


p(Dear | N) = 0.47
p(Friend | N) = 0.29
p(Lunch | N) = 0.18
p(Money | N) = 0.06

p(N) = 0.67

...and the probability that the word **Friend** occurs in **spam**.

$$p(\mathbf{S}) \times p(\mathbf{Dear} \mid \mathbf{S}) \times p(\mathbf{Friend} \mid \mathbf{S})$$

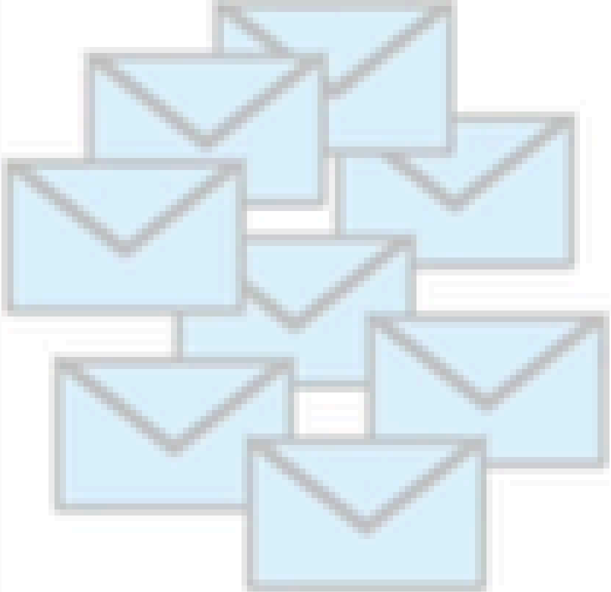


p(Dear | S) = 0.29
p(Friend | S) = 0.14
p(Lunch | S) = 0.00
p(Money | S) = 0.57

p(S) = 0.33

p(N | Dear Friend) ∝ 0.09 Dear Friend

Like before, we can think of **0.01** as the score that **Dear Friend** gets if it is Spam.



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

p(N) = 0.67

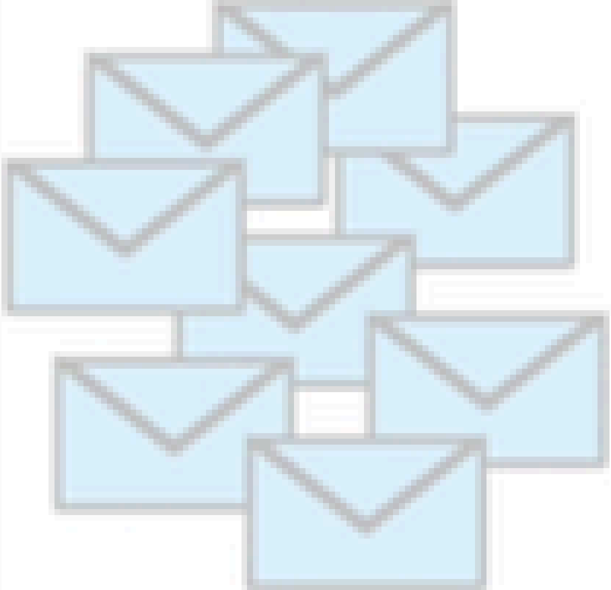
$$0.33 \times 0.29 \times 0.14 \boxed{= 0.01}$$

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

p(S) = 0.33

Dear Friend

Now that we understand the basics of how **Naive Bayes Classification** works...



$p(\text{Dear} \mid N) = 0.47$

$p(\text{Friend} \mid N) = 0.29$

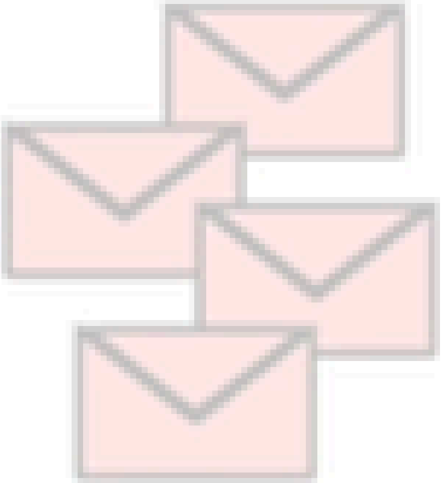
$p(\text{Lunch} \mid N) = 0.18$

$p(\text{Money} \mid N) = 0.06$

$p(N) = 0.67$

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

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



$p(\text{Dear} \mid S) = 0.29$

$p(\text{Friend} \mid S) = 0.14$

$p(\text{Lunch} \mid S) = 0.00$

$p(\text{Money} \mid S) = 0.57$

$p(S) = 0.33$