Example of a Bayes Theorem for Spam Filtering for a Simple Example

(http://hernan.amiune.com/teaching/from-a-bayes-theorem-naive-example-to-a-naive-bayes-example.html

The problem

You have a database of 100 emails.

- 60 of those 100 emails are spam
  - 48 of those 60 emails that are spam have the word "buy"
  - 12 of those 60 emails that are spam don't have the word "buy"
- 40 of those 100 emails aren't spam
  - 4 of those 40 emails that aren't spam have the word "buy"
  - 36 of those 40 emails that aren't spam don't have the word "buy"

What is the probability that an email is spam if it has the word "buy"?

Reformulate the previous problem by using probabilities instead of quantities.

The problem

You have a database of emails.

- 60% of those emails are spam
  - 80% of those emails that are spam have the word "buy"
  - 20% of those emails that are spam don't have the word "buy"
- 40% of those emails aren't spam
  - 10% of those emails that aren't spam have the word "buy"
  - 90% of those emails that aren't spam don't have the word "buy"

Again: What is the probability that an email is spam if it has the word "buy"?

The answer in Bayes Notation

First let's define some notation:
P(spam) = the probability that an email is spam
P(not spam) = the probability that an email isn't spam
P("buy"|spam) = the probability that an email that it is spam has the word "buy"
P("buy"|not spam) = the probability that an email that it isn't spam has the word "buy"
P(spam|"buy") = the probability that an email that has the word "buy" is spam

So P(spam|"buy") is the answer we are looking for.

P("buy"|spam) * P(spam) counts all the emails that are spam and have the word "buy"
P("buy"|not spam) * P(not spam) counts all the emails that aren't spam and have the word "buy"
Summing the previous two P("buy"|spam) * P(spam) + P("buy"|not spam) * P(not spam) we count all the emails that have the word "buy"

So our answer will be:

P(spam|"buy") = P("buy"|spam) * P(spam) / (P("buy"|spam) * P(spam) + P("buy"|not spam) * P(not spam))

And this is the equation known as Bayes Theorem

We end doing the same thing we did before. 48 / (48 + 4) but with probabilities 0.8 * 0.6 / (0.8*0.6 + 0.1*0.4) = 0.48 / 0.52