1. A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase “free money” is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention “free money”. What is the probability that it is spam?

Sol)

According to Bayes' theorem:

P(S | M) = (P(M | S) \* P(S)) / P(M)

P(M | S) is the probability that the email mentions "free money" given that it is spam, which is 0.1 (10%).

P(S) is the probability that any given email is spam, which is 0.8 (80%).

P(M) is the probability that the email mentions "free money", which can be calculated using the law of total probability:

P(M) = P(M | S) \* P(S) + P(M | non-S) \* P(non-S)

P(M | non-S) is the probability that the email mentions "free money" given that it is not spam, which is 0.01 (1%).

P(non-S) is the probability that any given email is not spam, which is 0.2 (20%).

Now we can substitute the values into the equation:

P(M) = (0.1 \* 0.8) + (0.01 \* 0.2) = 0.081

Finally, we can calculate P(S | M):

P(S | M) = (0.1 \* 0.8) / 0.081 ≈ 0.9877

Therefore, the probability that the email is spam given that it mentions "free money" is approximately 0.9877, or 98.77%.