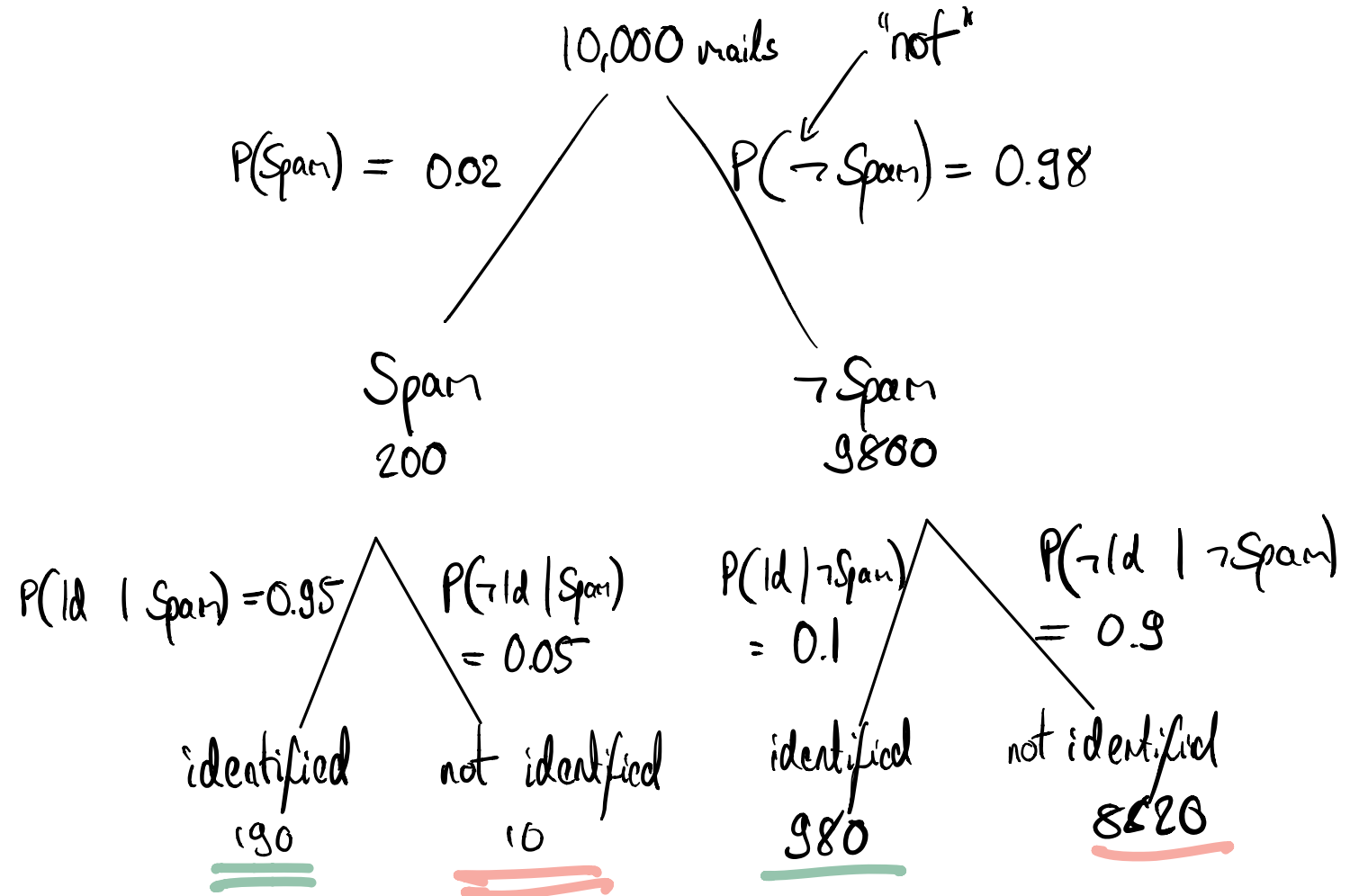


The rate of spam mails for a certain email address is 2%. A spam filter identifies a spam mail with a probability of 95%. At the same time, 10% of non-spam messages are classified as spam.

(a) What is the probability that a mail that was marked as spam is truly a spam mail?

(b) What is the probability that a mail that was not identified as spam is spam?



a)
$$P(\text{Spam} | \text{Id}) = \frac{190}{980 + 190} = \frac{190}{1170} \approx 0.162$$

$$= \frac{10,000 P(S) P(\text{Id} | S)}{10,000 P(S) P(\text{Id} | S) + 10,000 P(\neg S) P(\text{Id} | \neg S)}$$

Only about 1 in 6 mails classified as spam are actually spam! Why? High probability of non-spam messages (98%) and relatively high probability (10%) of a false positive.

b)
$$P(\text{Spam} | \neg \text{Id}) = \frac{10}{10 + 8820} = \frac{10}{8830} \approx 0.00113$$