Part 1 - Understanding Bayes' Theorem

Bayes' Theorem:

P(A | B) = (P(B | A) * P(A)) / P(B)

Problem Scenario:

Students often call in sick more frequently during holiday weeks.

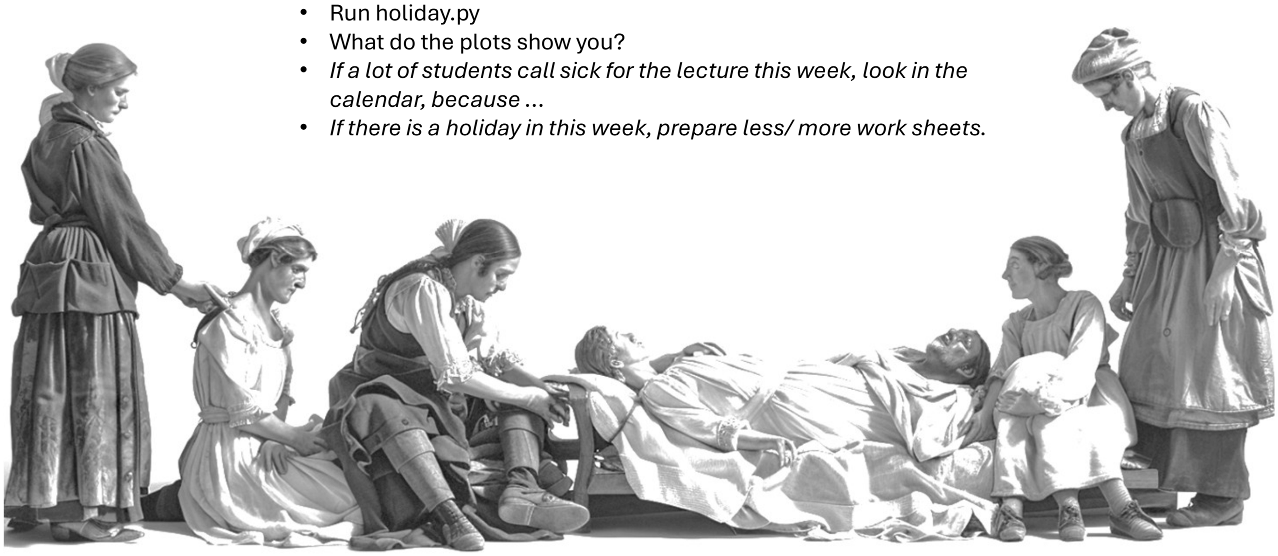
Assume:

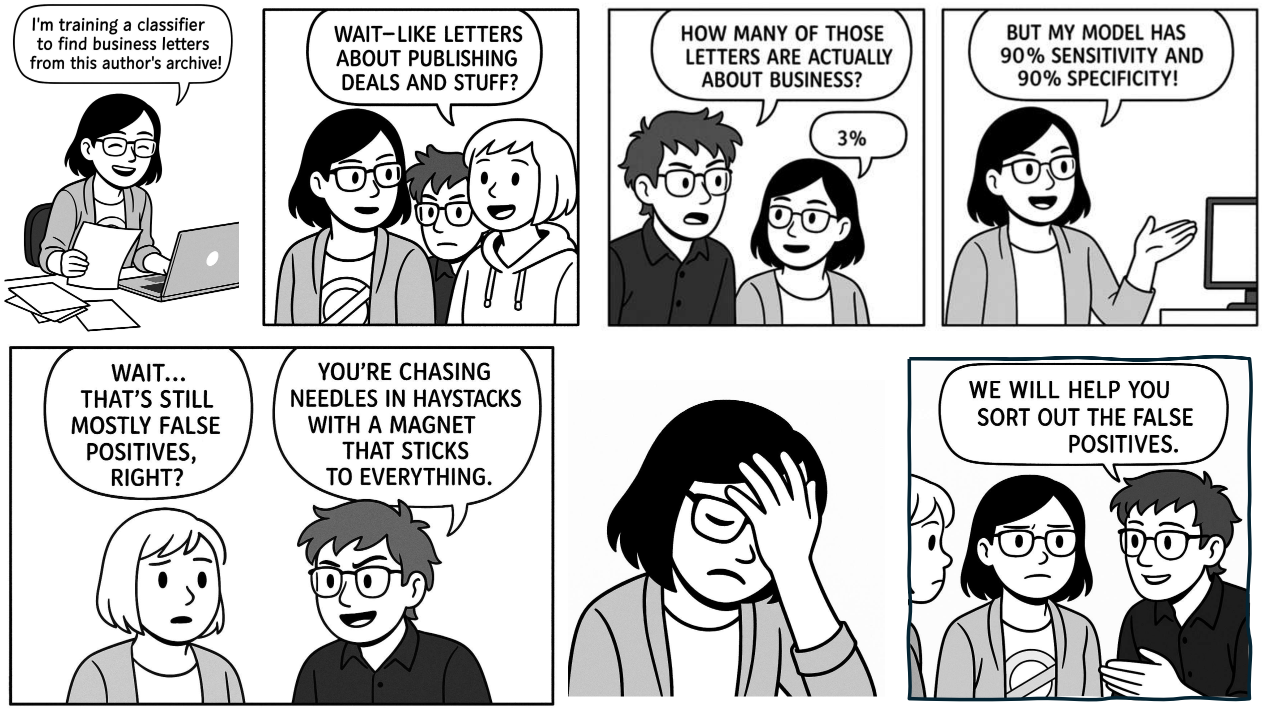
- P(Sick) = 0.1 (10% chance that a student is sick in any week)
- P(Holiday) = 0.2 (20% of weeks have a holiday)
- P(Sick | Holiday) = 0.25 (students are more likely sick in holiday weeks)

- What is the probability that a student was sick AND the week was a holiday?
  Hint: This is P(Sick and Holiday) = P(Sick | Holiday) * P(Holiday)
- What is the probability that a week had a holiday, given that a student was sick?
  Hint: Apply Bayes' Theorem. Write Python code similiar to below:

```
sick = 0.1
holiday = 0.2
sick_given_holiday = 0.25
holiday_given_sick = (sick_given_holiday * holiday) / sick
print(holiday_given_sick)
```

- Run holiday.py
- What do the plots show you?
- *If a lot of students call sick for the lecture this week, look in the calendar, because …*
- *If there is a holiday in this week, prepare less/ more work sheets.*





Run the script baseRateFallacy.py!

- How many “business letters” will Anna find in 1000 letters from the archive?
- How many of those are actually about business matters?
- What is the percentage of false positives?
- What is the difference between sensitivity and specificity?

Look in the code:

- How much better would be the results, if a better classifier was used?
- Try out 95% accuracy, 99% accuracy.
- Try out a larger letter corpus (10.000 letters).

During an archaeological excavation, a few Roman bricks are found at a site.
We consider two competing hypotheses:

- **H_1:** There was a permanent Roman settlement at this site.
- **H_2:** It was only a temporary station (e.g., a military camp or stopover).

Before any excavations, historians estimated (based on ancient travel reports) that:

- **$P(H_1)$ = 20% (0.2)**
- **$P(H_2)$ = 80% (0.8)**

New Evidence (E_1): Roman Coins

In a subsequent excavation, several Roman coins are discovered, minted under Emperor Hadrian.

We assume:

- **$P(E_1 \mid H_1)$ = 0.8** (high probability of coins if there was a permanent settlement)
- **$P(E_1 \mid H_2)$ = 0.2** (lower probability if it was only a temporary station)

$$P(E_1) = 0.8 \cdot 0.2 + 0.2 \cdot 0.8 = 0.32$$

$$P(H_1|E_1) = \frac{P(H_1|E_1) \cdot P(H_1)}{P(E_1)} \qquad P(H_1|E_1) = \frac{0.8 \cdot 0.2}{0.32} = 0.5$$

After finding the coins, the probability of a permanent settlement rises from 20% to 50%.

Later, remains of Roman buildings are uncovered.

Assumptions:

- **$P(E_2 \mid H_1)$ = 0.9** (very likely with a settlement)
- **$P(E_2 \mid H_2)$ = 0.1** (very unlikely with only a station)

We update using the previous posterior as the new prior:

- **$P(H_1)$ = 0.5**
- **$P(H_2)$ = 0.5**

After the discovery of building foundations, the probability of a permanent settlement increases to 90%.

A new artifact is found:

A small amulet dedicated to the god Hermes (god of travelers), often associated with temporary stations or traveler shelters.

Assume:

- **$P(E_3 \mid H_1)$ = 0.1** (less likely if there was a permanent settlement)
- **$P(E_3 \mid H_2)$ = 0.6** (more likely for a temporary station)
- Use Bayes' Theorem to update the current belief (after the last update where **$P(H_1)$ = 0.9**) in light of this new evidence.
- How would the conclusions change if the initial prior (before any evidence) had been **$P(H_1)$ = 0.5** instead of 0.2?
- Why is it important to carefully choose priors in Bayesian reasoning?