Naïve Bayesians

Back to Basics Series

09 Jan 2021

Goal

Developing the Bayesian muscle to solve a wide range of problems

Naïve Bayesian Philosophy

Intuitive (Visual) Understanding of the Bayesian Reasoning

Ability to model real world problems in a Bayesian Setting

Starting from Simple Probabilistic modelling

Adapting it in a a Bayesian setting
And moving towards ML models

Fluency in the Calculus of Bayesian Stats & ML model

PyMC3

Season 2: Back to Basics

Ep 1	Bayes Theorem
Ep 2 – Ep 3	Problems with Binomial Likelihoods
Ep 4	Disease Detection
Ep 5	Naive Bayes Classification
Ep 6	Gaussian Naive Bayes Classification
Ep 7	German Tank Problem
Ep 8	Waiting Times (Continuous Distributions)

Back to Basics

		Canonical Problem	Applications
Ep 1	Bayes Theorem	There are 2 boxes from which cookies can be taken from. Box A and Box B. Box A contains 10 chocolate cookies, Box B contains 5 ginger cookies. Given that you get a chocolate cookie which box was it taken from?	The Shy Librarian Problem Naive Bayes algorithm
Ep 2 Ep 3	Problems with Binomial Likelihoods	You have 2 coins C1 and C2. p(heads for C1) = .7 & P(heads for C2) = 0.6 You flip the coin 10 times. What is the probability that the given coin you picked is C1 given you have 7 heads and 3 tails?	A/B Testing
Ep 4	Disease Detection	A particular disease affects 1% of the population. There is an imperfect test for this disease: The test gives a positive result for 90% of people who have the disease, and 5% of the people who are disease-free. Given a positive test result – what is the probability of having the disease?	COVID Tests (PCR & Antibody)! Fraud Detection
Ep 5	Naive Bayes Classification	Given these words occur in this text what's the probability it's spam?	Any Classification Problem
Ep 6	Gaussian Naive Bayes Classification	Given the weights and heights of basketball players, what's the probability that person a is a basketball player given weight = w and height = h?	

Back to Basics

		Canonical Problem	Applications
Ep 7	German Tank Problem	Suppose tanks were given a serial number based on the order in which they were manufactured. Given that you've observed a tank with serial number "10", how many tanks were actually manufactured in total?	?
Ep 8	Waiting Times (Continuous Distributions)	Suppose you need to gather 10 patients for a trial. Each signup happens at time t_i (i=1, 10). How long do you have to wait after it took you 3 weeks to accrue 2 signups?	Planning Trials Estimating Queues

Bayes Rule

Posterior   Likelihood   Prior

$$P(\theta_i \mid D) = \frac{P(D \mid \theta_i)\, P(\theta_i)}{P(D)}$$

Normalising Constant

Bayes Rule

Posterior Likelihood Prior

$$P(\theta_i \mid D) = \frac{P(D \mid \theta_i)\, P(\theta_i)}{\sum_{all\ j} P(D \mid \theta_j)\, P(\theta_j)}$$

Normalising Constant

Canonical Problem

You have 2 coins C_1 and C_2.

$p(\text{heads for } C_1) = 0.7$ $p(\text{heads for } C_2) = 0.6$

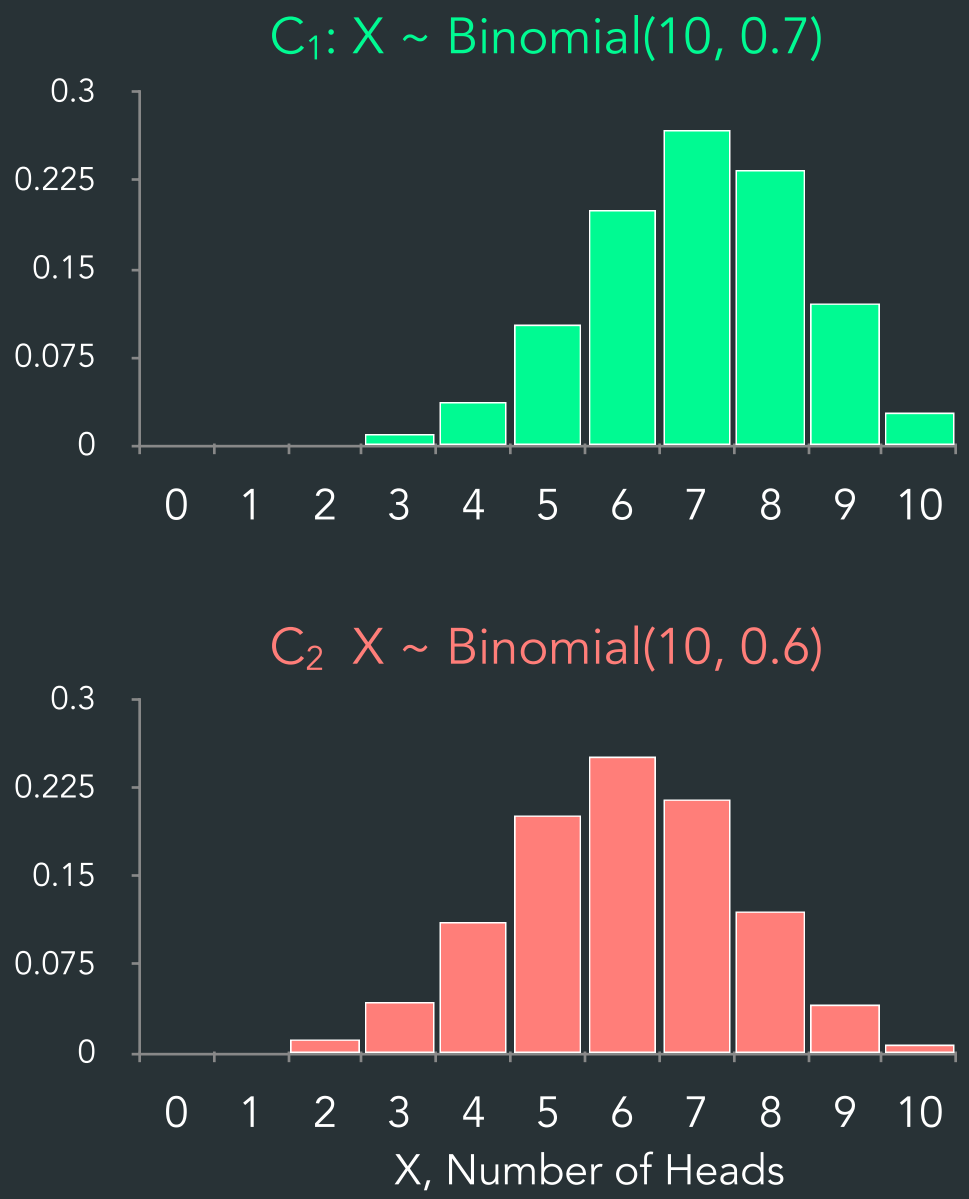
You flip one of the coins 10 times and get **7 heads** and **3 tails**

What is the probability that the given coin you picked is C_1?

What is the probability that the given coin you picked is C_1?

- $X \sim \text{Binomial}(n, p)$
  - $0 \leq X \leq n$
  - $n > 0$
  - $0 < p < 1$

$$P(X = k \mid p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

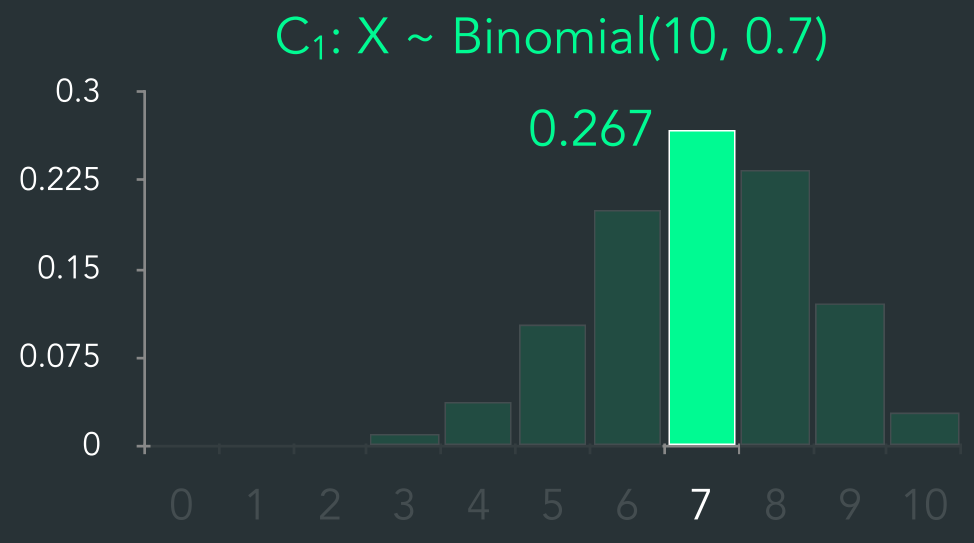


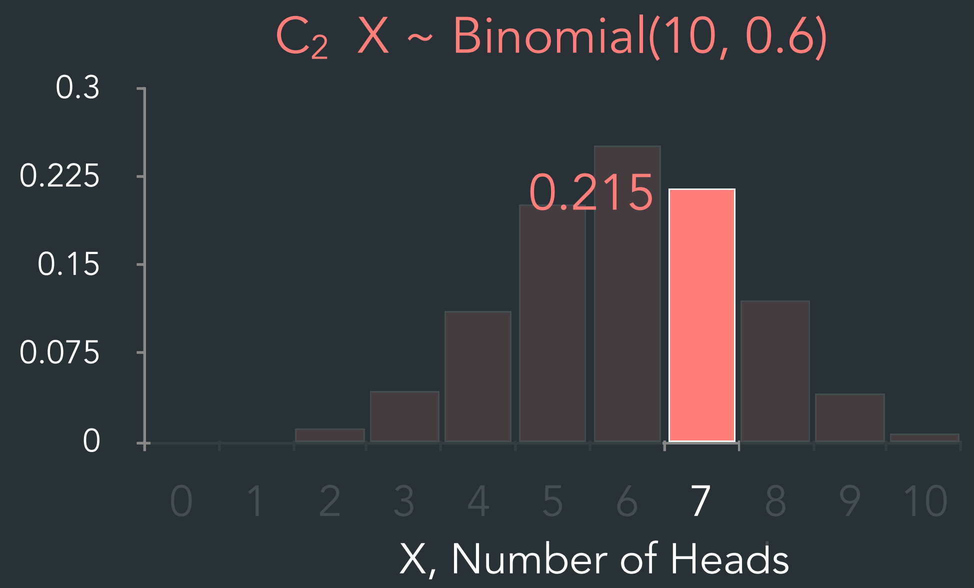
What is the probability that the given coin you picked is C_1?

- $X \sim \text{Binomial}(n, p)$
  - $0 \leq X \leq n$
  - $n > 0$
  - $0 < p < 1$

$$P(X = k \mid p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Likelihood

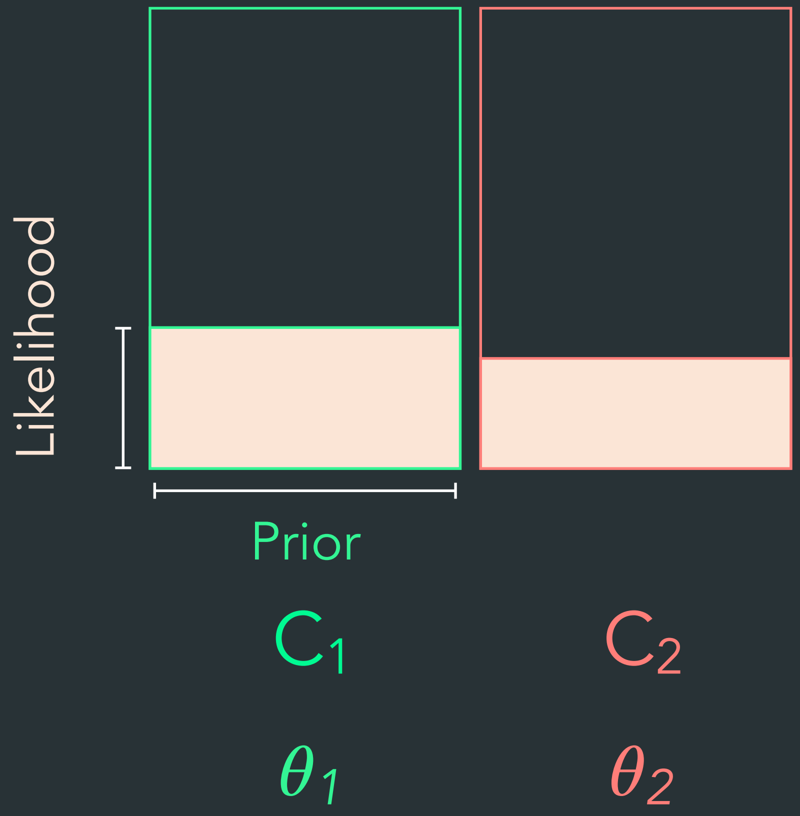




What's the probability a chocolate chip cookie came from box A?

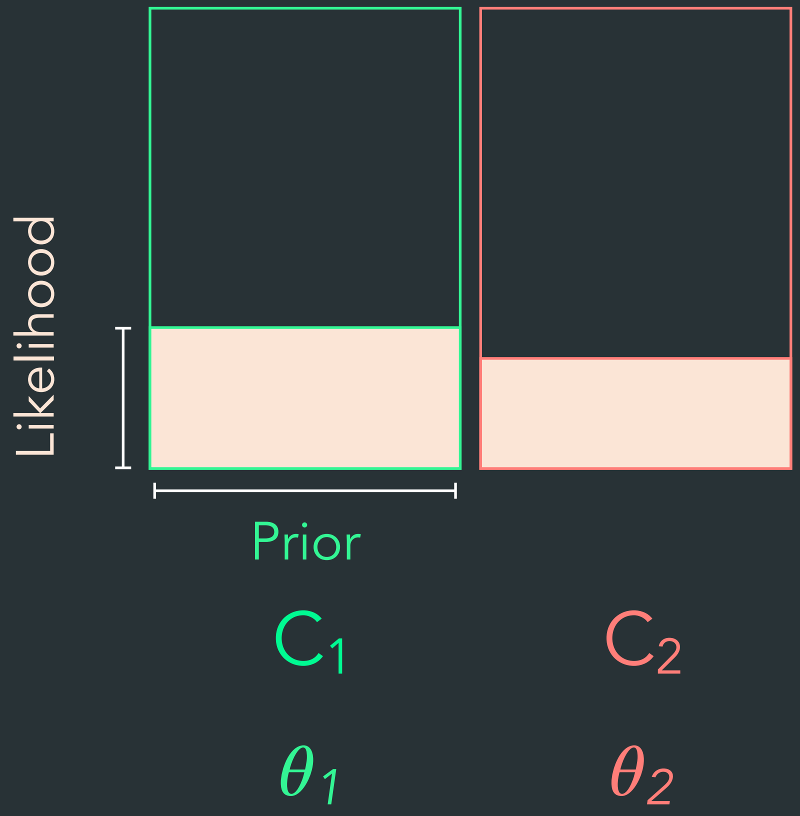
$$P(\theta_i \mid D) = \frac{P(D \mid \theta_i)\, P(\theta_i)}{\sum_{all\ j} P(D \mid \theta_j)\, P(\theta_j)}$$

What's the probability a chocolate chip cookie came from box A?



$$P(\theta_i \mid D) = \frac{P(D \mid \theta_i)\, P(\theta_i)}{\sum_{all\ j} P(D \mid \theta_j)\, P(\theta_j)}$$

What's the probability a chocolate chip cookie came from box A?



$$P(\theta_i \mid D) = \frac{P(D \mid \theta_i)\, P(\theta_i)}{\sum_{all\ j} P(D \mid \theta_j)\, P(\theta_j)}$$

$$P(C_1 \mid X=7)$$

$$= \frac{P(X=7 \mid C_1)\, P(C_1)}{P(X=7 \mid C_1)\, P(C_1) + P(X=7 \mid C_2)\, P(C_2)}$$

What’s the probability a chocolate chip cookie came from box A?

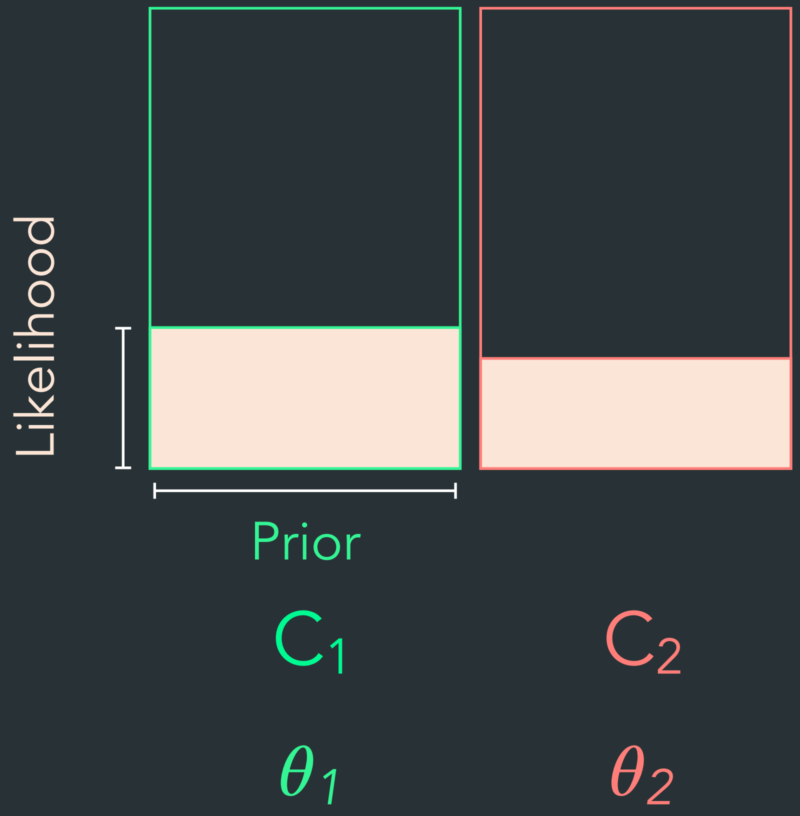
C_1 $\quad$ C_2

θ_1 $\quad$ θ_2

$$P(C_1 \mid X=7)$$

$$= \frac{P(X=7 \mid C_1)\, P(C_1)}{P(X=7 \mid C_1)\, P(C_1) + P(X=7 \mid C_2)\, P(C_2)}$$

What's the probability a chocolate chip cookie came from box A?

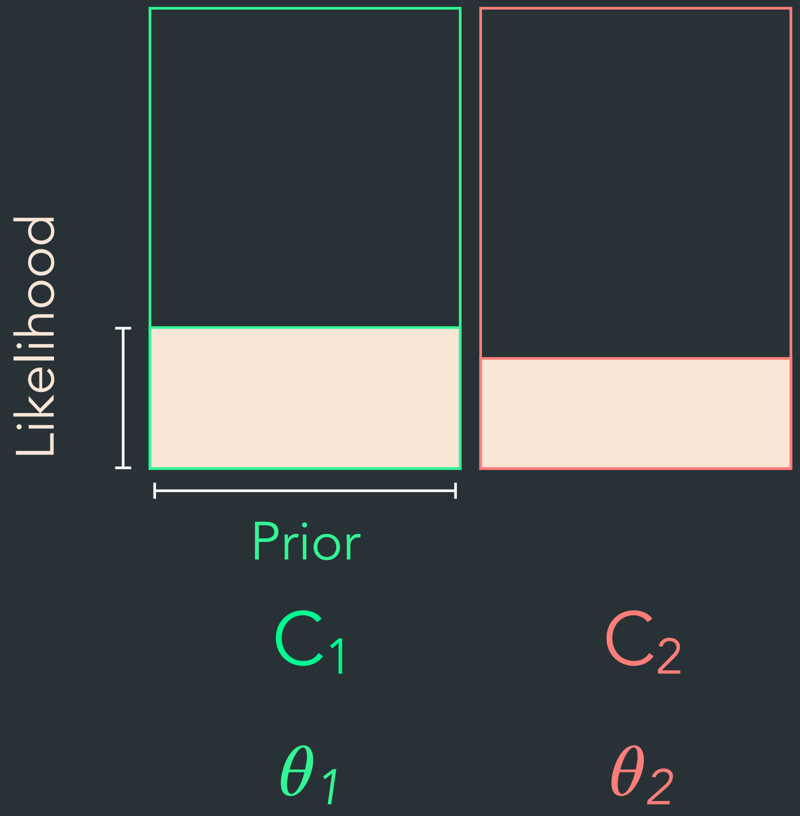


$$P(C_1 \mid X=7)$$

$$= \frac{P(X=7 \mid C_1)\, P(C_1)}{P(X=7 \mid C_1)\, P(C_1) + P(X=7 \mid C_2)\, P(C_2)}$$

$$= \frac{0.267\, P(C_1)}{0.267\, P(C_1) + 0.215\, P(C_2)}$$

What's the probability a chocolate chip cookie came from box A?



$$P(C_1 \mid X=7)$$

$$= \frac{P(X=7 \mid C_1)\, P(C_1)}{P(X=7 \mid C_1)\, P(C_1) + P(X=7 \mid C_2)\, P(C_2)}$$

$$= \frac{0.267\, P(C_1)}{0.267\, P(C_1) + 0.215\, P(C_2)}$$

$$= \frac{0.267 \times 0.5}{0.267 \times 0.5 + 0.215 \times 0.5}$$

$$= 0.554$$

Canonical Problem v2

You randomly draw a coin from 100 coins

$C_{u:}$ 1 unfair coin (head-head) p(heads for C_u) = 1

$C_{f:}$ 99 fair coins (head-tail) p(heads for C_f) = 0.5

You flip it 10 times and the result is 10 heads

Whats the probability that the coin is unfair?

Whats the probability that the coin is unfair?

$P(C_u) = 1/100$ $P(C_f) = 99/100$

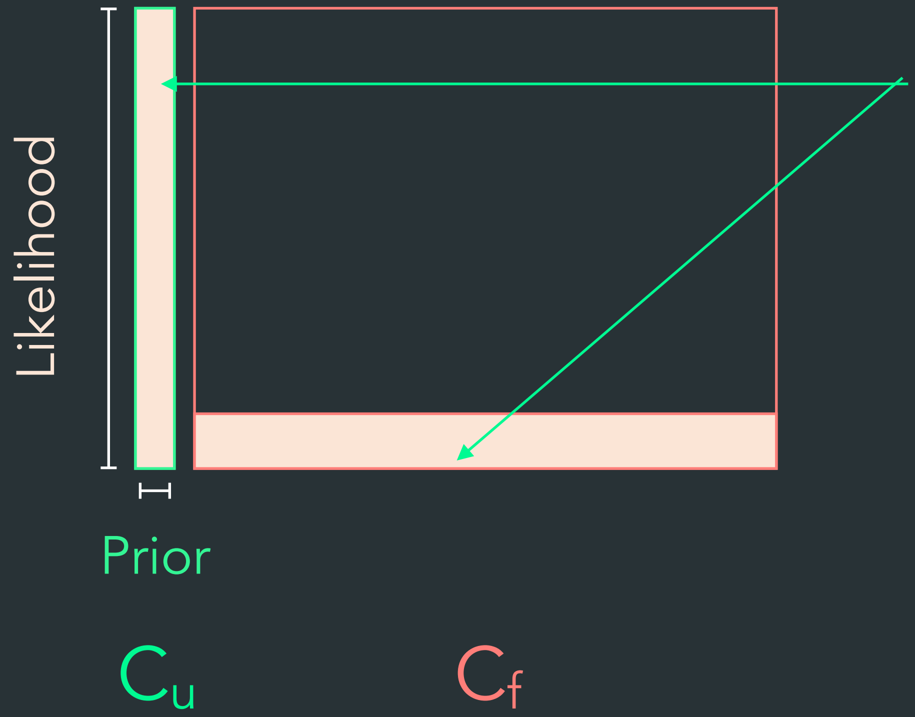
Whats the probability that the coin is unfair?

$$P(C_u \mid X=10)$$

$$= \frac{P(X=10 \mid C_1)\, P(C_u)}{P(X=10 \mid C_1)\, P(C_u) + P(X=10 \mid C_f)\, P(C_f)}$$

$P(C_u) = 1/100$ $\quad$ $P(C_f) = 99/100$

Whats the probability that the coin is unfair?



Same "Area"

Bernoulli Distribution

- Y ~ Bernoulli(p)
  - X = 0, 1
  - 0 < p < 1
- Events with 1 Trial & 2 Possible Outcomes
- Examples
  - 1 coin flip
  - Answers to True/False Quiz
  - Voting in a 2-Candidate Election
  - Result of a COVID-19 Test