

# Naive Bayes



## Por que?

Processamento que usa as informações que temos a nosso favor. Ou seja, se sabemos o % de pessoas que estão indo pescar em dia de chuva, prevemos quantas pessoas vão em dias chuvosos em nosso pesque pague.

## Matematicamente como acontece?

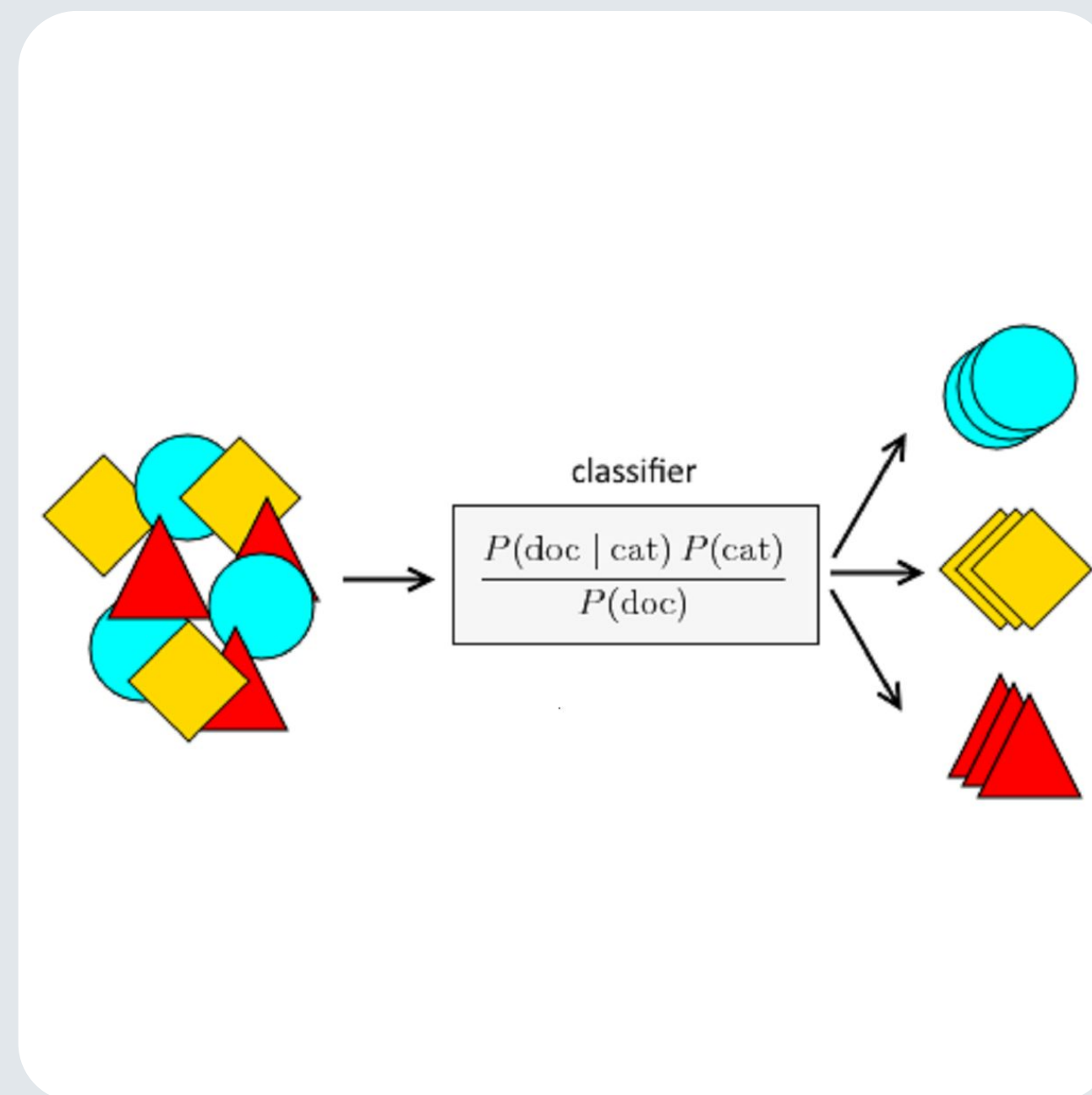
Fazemos uma matriz que considera a chance percentual de algo acontecer dado alguns cenários (conjunto de variáveis).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Diagram illustrating the components of the equation:

- $P(c|x)$  is labeled "Probabilidade posterior" (Posterior Probability).
- $P(x|c)$  is labeled "Probabilidade original da Classe" (Original Class Probability).
- $P(c)$  is labeled "Probabilidade" (Probability).
- $P(x)$  is labeled "Preditor da probabilidade posterior" (Posterior Probability Predictor).

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$





## Vantagens e desvantagens

As vantagens deste algoritmo é ter um treinamento rápido, insensível a dados irrelevantes (os ignora) e também tem bons resultados com dados discretos e contínuos.

E uma das desvantagens é que não existe uma independência de características. Por isso o nome naive - 'ingenuo'.

## Na prática

Na prática, a biblioteca Scikit Learn possui modelos prontos para serem treinados, possuindo uma documentação bem escrita para estudo.