Naive Bayes

Supervised Algorithm

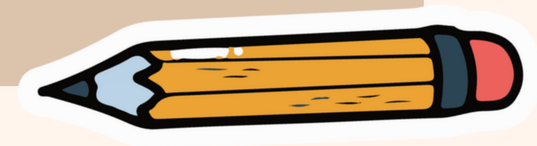
START

Background

Naive Bayes is classification algorithm based on Bayes' Theorem.

It assumes that all the features that predict the target value are independent of each other.

It works with huge dataset and mostly used to solve text kinds of data.

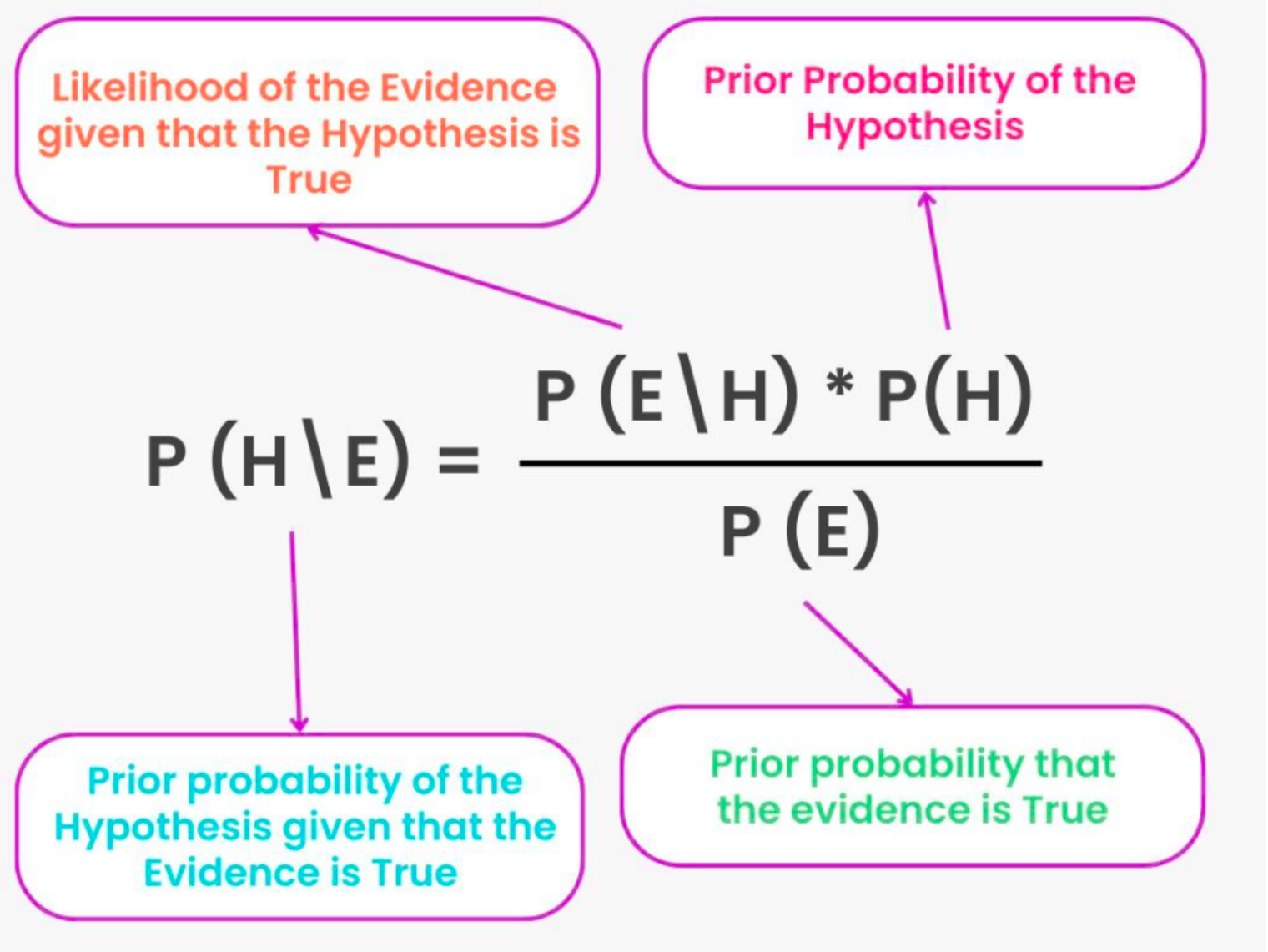


What it does?

1. Calculates probability of each class

2. Picks one with highest probability

Bayes Theorem



Problem Statement

Patient's probability of having liver disease if they are an alcoholic.

If ‘A’ is the event “Patient has liver disease.” Past data tells you that 10% of patients entering your clinic have liver disease.

$P(A) = 0.10$

‘B’ could mean that “Patient is an alcoholic.” 5% of the clinic’s patients are alcoholics.

$P(B) = 0.05$

If you know that among those patients diagnosed with liver disease, 7% are alcoholics.

This is your ‘B|A’: the probability that a patient is alcoholic, given that they have liver disease, is 7%.

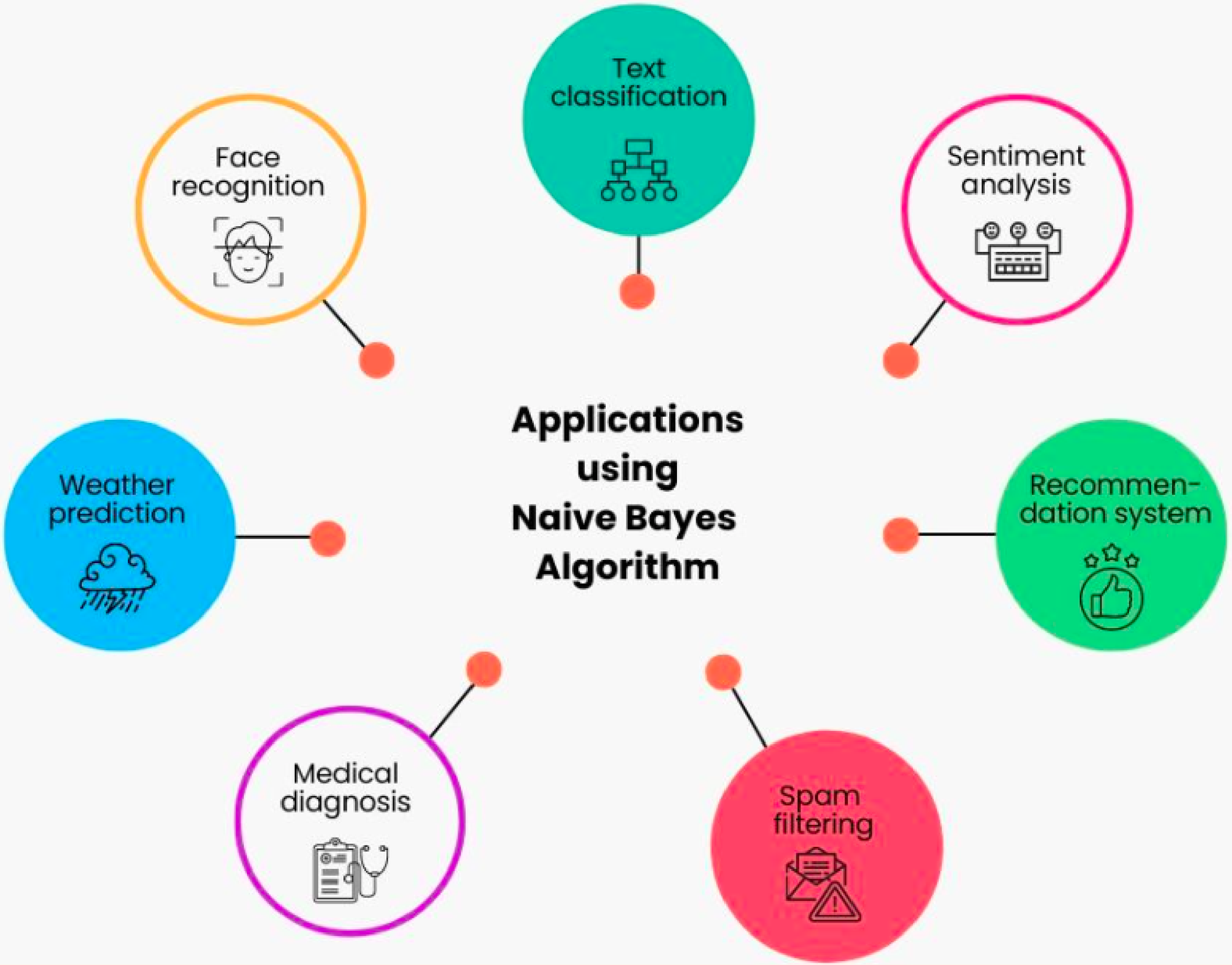
$P(B|A) = 0.05$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' theorem tells you:

P(A|B) = (0.07 * 0.1)/0.05 = 0.14

In other words, if the patient is an alcoholic, their chances of having liver disease is 0.14 (14%).


Text classification
Face recognition
Sentiment analysis
Applications using Naive Bayes Algorithm
Weather prediction
Recommen-dation system
Medical diagnosis
Spam filtering

TYPES

- Gaussian Naive Bayes
- Bernoulli Naive Bayes
- Multinominal Naive Bayes
- Optimal Naive Bayes

Advantages

- Doesn’t require larger amounts of training data
- Straightforward to implement
- Convergence is quicker than other models
- Highly scalable
- Can handle both continuous and categorical data