Naïve Bayes Algorithm

28 October 2025 10:58

Naïve Bayes is a **probabilistic classification algorithm** based on **Bayes' theorem** — it predicts the class of a sample by computing probabilities of different possible classes given the input features.

It's called **"naïve"** because it **assumes independence** among features — meaning each feature contributes to the final decision **independently**.

Despite this strong (and unrealistic) assumption, it performs **surprisingly well** in many real-world applications like:

- Spam filtering
- Sentiment analysis
- Document classification
- Medical diagnosis

Naïve Bayes: Mathematical concepts

28 October 2025 10:33

🧠 1. What is Probability?

Probability measures **how likely** an event is to happen.

For an event A, the probability is:

$$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

What is an Event in Probability?

An **event** is a specific outcome or a set of outcomes of a random experiment.

- **Sample Space (S):**

  The set of all possible outcomes.

  Example: For a coin toss → S={H,T}

- **Event (A):**

  A subset of the sample space.

  Example: Event A = "getting a head" → A={H}

Types of Events:

Independent Events:

Two events A and B are said to be independent if the occurrence of one does not affect the probability of occurrence of the other.

Formally:

$$P(A \cap B) = P(A) \cdot P(B)$$

or equivalently,

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

🎲 Example 1 — Coin Toss:

- Event A: Get a head on the first toss → $P(A) = \frac{1}{2}$
- Event B: Get a head on the second toss → $P(B) = \frac{1}{2}$

Both coin tosses are independent — what happens in the first doesn't affect the second.

So:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

✅ The result matches the intuition — both heads occur 1/4 of the time.

🎲 Example 2 — Dice Roll:

- Event A: Getting an even number → $P(A) = \frac{3}{6} = 0.5$
- Event B: Getting a number greater than 4 → $P(B) = \frac{2}{6} = \frac{1}{3}$

Now check independence:

$$P(A \cap B) = P(\text{Even and } {>}4) = P(\{6\}) = \frac{1}{6}$$

$$P(A)P(B) = 0.5 \times \frac{1}{3} = \frac{1}{6}$$

✅ Since $P(A \cap B) = P(A)P(B)$, the two events are **independent**.

Dependent Events:

📘 Definition:

Two events **A** and **B** are **dependent** if the occurrence of one **affects** the probability of occurrence of the other.

Formally:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

and $P(B|A) \neq P(B)$.

🧩 Intuition:

If events are dependent, knowing that one event happened **changes** the probability of the other.

🃏 Example 1 — Drawing Cards *Without Replacement*:

- A deck has 52 cards.
- Event A: Drawing a king on the first draw → $P(A) = \frac{4}{52} = \frac{1}{13}$
- Event B: Drawing a king on the second draw **without replacement**.

If one king was already removed:

$$P(B|A) = \frac{3}{51}$$

So:

$$P(A \cap B) = P(A) \times P(B|A) = \frac{1}{13} \times \frac{3}{51} = \frac{3}{663}$$

✅ Here, B's probability **depends** on whether A occurred — hence **dependent events**.

🧮 Example 2 — Selecting Balls:

A box contains 3 red and 2 blue balls.

- Event A: First ball drawn is red.
- Event B: Second ball drawn is red.

If the first red ball is not replaced:

$$P(A) = \frac{3}{5}, \quad P(B|A) = \frac{2}{4} = \frac{1}{2}$$
$$P(A \cap B) = \frac{3}{5} \times \frac{1}{2} = \frac{3}{10}$$

Since $P(B|A) \neq P(B)$, events A and B are **dependent**.

🎯 2. Conditional Probability

Definition:

Conditional probability measures the chance of an event A occurring **given that** another event B has already occurred.

It's written as:

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

provided $P(B) > 0$.

🧩 Intuitive Meaning:

- $P(A)$: Probability of A happening.
- $P(A|B)$: Probability of A happening **when we already know** B happened.
- So, it "conditions" on some known information.

3. Bayes' Theorem

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

where:

- $P(A|B)$: Posterior — probability of A after observing B
- $P(B|A)$: Likelihood — probability of B given A
- $P(A)$: Prior — initial belief about A
- $P(B)$: Evidence — overall probability of observing B

Naive Bays theorm:

Dataset

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

	x_1	x_2	x_3 ... x_n		Y
1	7.5	1	3	2	y
2					N
...					y
100					N

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

$$P(Y|x_1, x_2, x_3 \cdot x_n) = \frac{[P(x_1/y) \cdot P(x_2/y) \cdot P(x_3/y) \ldots P(x_n/y)]P(y)}{P(x_1) \cdot P(x_2) -- P(x_n)}$$

constant

1

$$P(N|x_1 x_2 x_3 \cdots x_n) = \frac{[P(x_1/N)\ P(x_2/N) -- P(x_n/N)]\ P(N)}{P(x_1) \cdot P(x_2) --- P(x_n)}$$

constant

2

$$P(y|x_1 - x_n) = [P(x_1/y)\ P(x_2/y) \cdots P(x_n/y)] \cdot P(y)$$

$$P(N|x_1, x_2 \cdots x_n) = [P(x_1/N)\ P(x_2/N) --- P(x_n/N)] \cdot P(N)$$

Example

28 October 2025 11:04

$$P(y|x_1 \ldots x_n) = [P(x_1/y)\,P(x_2/y) \ldots P(x_n/y)] \cdot P(y)$$

$$P(N|x_1, x_2 \ldots x_n) = [P(x_1/N)\,P(x_2/N) \ldots P(x_n/N)] \cdot P(N)$$

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes ✓
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No ✓
Rainy	No ✓
Sunny	Yes
Rainy	Yes ✓
Sunny	No ✓
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency table

	Yes	NO	P(y)	P(N)
Sunny	3	2	$P(\text{Sunny}/y) = 3/5$	2/5
Overcast	~~4~~ 3	0	4/4	0
Rainy	2	3	2/5	3/5
	9	5		

$= 14$ $\quad P(\text{yes}) = 9/14, \quad P(N) = 5/14$

$$P(\text{Yes}|\text{sunny, overcast, Rainy}) = P(\text{Sunny}/\text{yes}) \cdot P(\text{overcast}/\text{yes}) \cdot P(\text{Rainy}/\text{yes}) \cdot P(\text{Yes})$$

$$= [3/5 \times 1 \times 2/5] \times \frac{9}{14}$$

$$= 3/5 \times 2/5 \times 9/14 = 0.154$$

$$P(\text{NO}|\text{Sunny, overcast, Rainy}) = P(\text{Sunny}/\text{NO})\,P(\text{Over}/\text{NO})\,P(\text{Rainy}/\text{No})\,P(\text{No})$$

$$= [2/5 \times 0 \times 3/5] \times \frac{5}{14} = 0$$

$$P(Y)_{\text{Normalize}} = \frac{P(\text{Yes})}{P(\text{yes}) + P(\text{No})} = \frac{0.154}{0.154 + 0} = 1$$

$$P(N)_{\text{Norm}} = \frac{P(N)}{P(N) + P(Y)} = \frac{0}{0.154 + 0} = 0$$

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

	Y	N
sunny	2	3
Hot	2	2

$P(\text{sunny}/y) = 2/5$, $P(\text{sun}/N) = 3/5$

$P(\text{Hot}/Y) = 2/4$ $P(\text{Hot}/\text{No}) = 2/4$

$P(\text{Yes}) = \frac{9}{14}$ $P(\text{No}) = \frac{5}{14}$

The player will play if day is sunny and hot?

$$P(Y/\text{sunny, hot}) = P(\text{sunny}/y) \cdot P(\text{Hot}/y)\ P(y)$$

$$= 2/5 \times \frac{2}{4} \times \frac{9}{14} = 0.128$$

$$P(N/\text{sunny, hot}) = P(\text{sunny}/N)\ P(\text{hot}/N) \cdot P(N)$$

$$= 3/5 \times \frac{2}{4} \times \frac{5}{14} = 0.107$$

$$P(Y/\text{sun, hot})_{\text{Norm}} = \frac{0.128}{0.128 + 0.107} = 0.544$$

$$P(N/\text{sunny, hot})_{\text{normalize}} = \frac{0.107}{0.128 + 0.107} = 0.456$$

Types of Naïve Bayes Classifiers

28 October 2025 11:02

Type	Data Type	Likelihood Model	Example Use
1. Gaussian Naïve Bayes	Continuous	Features assumed to follow a normal (Gaussian) distribution	Iris dataset, sensor data
2. Multinomial Naïve Bayes	Discrete (counts)	Features are frequencies or counts	Text classification, word counts
3. Bernoulli Naïve Bayes	Binary	Features are 0/1 indicators	Spam filtering, sentiment analysis