Applying Bayes' Theorem to Structured Data

Naive Bayes Classification Example with Manual Calculation

Syed Hamed Raza

UMT Lahore

December 2, 2025

Outline

Why Naive Bayes for Structured Data?

- Works on both categorical and numerical features.
- Extremely fast: closed-form probability estimation.
- Performs well even with small datasets.
- Useful for student performance prediction, risk detection, analytics.
- Clear and interpretable model for teaching probability-based ML.

Bayes' Rule for Classification

$$P(C \mid X) = \frac{P(X \mid C)\, P(C)}{P(X)}$$

- C: class label (Pass, Fail)
- X: feature vector (StudyTime, Attendance, Difficulty)
- Choose class with maximum posterior.

Naive Independence Assumption

$$P(X \mid C) = \prod_{i=1}^{n} P(x_i \mid C)$$

- Each feature is conditionally independent given class.
- Enables simple probability tables.
- Works well for tabular classification problems.

Training Dataset (Categorical Features)

ID	StudyTime	Attendance	Difficulty	Label
S1	Low	Poor	Hard	Fail
S2	Medium	Poor	Hard	Fail
S3	High	Good	Medium	Pass
S4	Medium	Good	Easy	Pass
S5	High	Good	Hard	Pass
S6	Low	Good	Medium	Fail

- Total samples: 6
- Pass: 3
- Fail: 3

Class Priors

$$P(Pass) = \frac{3}{6} = 0.5$$

$$P(Fail) = \frac{3}{6} = 0.5$$

Equal number of Pass and Fail instances.

Feature Likelihoods

We use Laplace smoothing:

$$P(x_i \mid C) = \frac{\text{count}(x_i, C) + 1}{N_C + k}$$

Where:

- N_C: total samples in class C
- k: number of possible values for the feature

StudyTime Likelihoods

StudyTime	**Pass**	**Fail**
Low	0	2
Medium	1	1
High	2	0

Denominator with smoothing:

$$P(x \mid Pass) = \frac{count + 1}{6} \quad P(x \mid Fail) = \frac{count + 1}{6}$$

Attendance Likelihoods

Attendance	Pass	Fail
Good	3	1
Poor	0	2

$$P(x \mid C) = \frac{count + 1}{5}$$

Two values: Good, Poor.

Difficulty Likelihoods

Difficulty	**Pass**	**Fail**
Easy	1	0
Medium	1	1
Hard	1	2

$$P(x \mid C) = \frac{count + 1}{6}$$

Feature has 3 possible values.

Test Instance

Predict label for:

- StudyTime = High
- Attendance = Good
- Difficulty = Hard

We compute:

$$P(Pass \mid X) \quad \text{and} \quad P(Fail \mid X)$$

Posterior for Pass

$$P(High \mid Pass) = \frac{2+1}{6} = 0.5$$

$$P(Good \mid Pass) = \frac{3+1}{5} = 0.8$$

$$P(Hard \mid Pass) = \frac{1+1}{6} = 0.333$$

$$P(Pass \mid X) \propto 0.5 \times 0.5 \times 0.8 \times 0.333$$

$$= 0.0666$$

Posterior for Fail

$$P(High \mid Fail) = \frac{0+1}{6} = 0.1667$$
$$P(Good \mid Fail) = \frac{1+1}{5} = 0.4$$
$$P(Hard \mid Fail) = \frac{2+1}{6} = 0.5$$

$$P(Fail \mid X) \propto 0.5 \times 0.1667 \times 0.4 \times 0.5$$
$$= 0.01667$$

Final Prediction

$$P(Pass \mid X) = 0.0666 \quad > \quad P(Fail \mid X) = 0.01667$$

Predicted Class

PASS

The test record is classified as **Pass** based on Naive Bayes.

Key Takeaways

- Naive Bayes easily applies to categorical structured data.
- Requires only simple frequency tables.
- Laplace smoothing handles zero-count problems.
- Provides interpretable probability-based decisions.