

Naive Bayes

Naive Bayes é um **algoritmo de classificação** baseado no **Teorema de Bayes**, que calcula a probabilidade de um evento ocorrer dado que outro evento já aconteceu.

É chamado de "**Naive**" porque **assume que todas as variáveis são independentes**, o que raramente é verdade na prática.

Ele calcula a probabilidade de um dado pertencer a uma classe, com base nas características fornecidas.

O objetivo é encontrar a **classe mais provável** para cada novo dado. A fórmula básica é:

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

Onde:

- $P(A|B)$: Probabilidade de ocorrer A dado que B ocorreu (classificação que queremos prever)
- $P(B|A)$: Probabilidade de B dado que A é verdadeiro
- $P(A)$: Probabilidade de A ocorrer (probabilidade a priori)
- $P(B)$: Probabilidade de B ocorrer

Naive Bayes é um modelo bem usado especialmente em **tarefas de processamento de linguagem natural (NLP)**, como:

- **Classificação de e-mails**
- **Análise de sentimentos**
- **Sistemas de recomendação simples**

Naive Bayes ainda é utilizado em cenários onde simplicidade, velocidade e facilidade de implementação são prioridades.

Abordagem Probabilística

Teorema de Bayes é muito utilizado para tomada de decisões na área de estatística, e esse algoritmo é baseado nesse teorema.

Exemplo de como fazer o teorema:

Primeiro você tem sua Base origem, com essa base, o objetivo é fazer uma classificação de probabilidade.

Exemplo de Base Origem

Base original				
História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

O primeiro passo é fazer a contagem do risco, pegando a classe que é o atributo que pretendemos fazer a previsão e a coluna previsora.

Lembrando que antes de contar é necessário fazer uma contagem total, no caso na tabela acima notamos que temos

- 6 Classes de risco Alto
- 3 Classes de risco Moderado
- 5 Classes de risco Baixo

Somando 14 Aparições no total

Exemplo:

- Nos casos em que história de crédito for boa, temos um risco considerado alto, ou seja (1/6)
- Nos casos em que história de crédito for boa temos 1 riscos moderados, ou seja (1/3)
- Nos casos em que história de crédito for boa temos 3 casos de risco baixo (3/5)

Com essas informações destacadas, começamos a montar o teorema de Bayes, conforme destacado abaixo:

