

## Pré - aula

Anápolis, 19 de abril de 2023.

Docente: Matheus Marques Portela

Nome da disciplina: Probabilidade e estatística

RA: 2310823

## RESPOSTA

Em aprendizado de máquina, há as seguintes tarefas: regressão e classificação. O algoritmo Naive Bayes é um classificador probabilístico baseado no teorema de Bayes.

Suponha a seguinte situação:

Você é proprietário de um clube esportivo e deseja saber qual é a probabilidade de os atletas praticarem ou não exercícios em dias nublados. Confira as informações a seguir:

Considerando os dados coletados, responda: Os atletas irão praticar esportes quando o tempo estiver nublado? Sim ou não? Justifique a sua resposta. Para isso, preencha as tabelas dos passos 1 e 2 e complemente os valores faltantes na expressão do teorema de Bayes.

tempo	não	sim
nublado	0	4
sol	2	3
chuva	3	2
<b>total</b>	<b>5</b>	<b>9</b>

$$\frac{4}{14} = 0,285 \quad \frac{5}{14} = 0,357$$

$$\frac{9}{14} = 0,642$$

ocorrência	Percentual
nublado	$0,285 \Rightarrow 28\%$
sol	$0,357 \Rightarrow 35\%$
chuva	$0,357 \Rightarrow 35\%$
sim	$0,642 \Rightarrow 64\%$
não	$0,357 \Rightarrow 35\%$

$$P(\text{sim} | \text{nublado}) = \frac{P(0,285 | 0,642) \cdot (0,642)}{P(0,285)}$$

$$P = \frac{0,3829 \cdot 0,642}{0,285} = 0,4120 \text{ ou } 41,20\%$$

Sim, pois a resposta com sim deu 41% e ele foi maior do que não.

Durante vários dias, você observou e coletou alguns dados do comportamento dos atletas. Na Tabela 1, a seguir, você inseriu as instâncias, que são as linhas dos registros e duas variáveis: Tempo (sol, nublado e chuva) e Praticar esportes (sim e não). Ou seja, você anotou se em dias de sol, chuva ou nublado, os atletas praticavam ou não exercícios.

Tempo	Praticar esportes
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Chuva	Sim
Sol	Sim
Nublado	Sim
Nublado	Sim
Chuva	Não

Considerando os dados da Tabela 1, você deseja fazer uma previsão para saber quantos atletas irão praticar esportes, com o tempo nublado. Ou seja, você deseja saber a probabilidade de um evento ocorrer (praticar esportes), com base na ocorrência de um outro evento (tempo).

Para chegar a esta resposta, será necessário seguir três passos. Veja a seguir quais são eles.

### 1º PASSO

Primeiramente, deve-se realizar a conversão do conjunto de dados em uma tabela de frequência, que consiste na contagem e na soma dos valores para cada instância aferida na coleta de dados.

Tempo	Não	Sim
Nublado		
Sol		
Chuva		
<b>Total</b>		

### 2º PASSO

Em seguida, deve-se criar uma tabela de probabilidade para encontrar a probabilidade de cada ocorrência e de cada combinação.

Ocorrência	Fórmula	Percentual
Nublado	$\frac{\text{Total de ocorrências de tempo nublado}}{\text{Total de ocorrências de todos os tempos}}$	
Sol	$\frac{\text{Total de ocorrências de tempo sol}}{\text{Total de ocorrências de todos os tempos}}$	
Chuva	$\frac{\text{Total de ocorrências de tempo chuva}}{\text{Total de ocorrências de todos os tempos}}$	
Sim	$\frac{\text{Total de ocorrências de praticar esportes sim}}{\text{Total de ocorrências de todos os resultados de praticar esportes}}$	
Não	$\frac{\text{Total de ocorrências de praticar esportes não}}{\text{Total de ocorrências de todos os resultados de praticar esportes}}$	

### 3º PASSO

Por fim, deve-se utilizar a equação do teorema de Bayes para calcular a probabilidade posterior para cada classe. A classe com maior probabilidade posterior é o resultado da previsão.

O teorema de Bayes é expresso matematicamente na seguinte equação:

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

- A e B são eventos e  $P(B) > 0$ .
- $P(A | B)$  são as probabilidades anteriores de A e B.
- $P(A | B)$  é a probabilidade posterior de A condicional a B.
- $P(B | A)$  é a probabilidade posterior de B condicional a A.

Dessa forma, para descobrir se os atletas irão praticar esportes quando estiver nublado, deve-se aplicar o teorema de Bayes da seguinte forma:

$$P(\text{Sim} | \text{Nublado}) = P(\text{Nublado} | \text{Sim}) \cdot P(\text{Sim}) / P(\text{Nublado})$$