

# Naive Bayes

Naive Bayes é um algoritmo baseado na questão da probabilidade, ou seja, vamos supor a seguinte base:

História do crédito	Dívida	Garantias	Renda Anual	Risco
Ruim	Alta	Nenhuma	<15000	Alto
Desconhecida	Alta	Nenhuma	>=15000 a <=35000	Alto
Desconhecida	Baixa	Nenhuma	>=15000 a <=35000	Moderado
Desconhecida	Baixa	Nenhuma	>35000	Alto
Desconhecida	Baixa	Nenhuma	>35000	Baixo
Desconhecida	Baixa	Adequada	>35000	Baixo
Ruim	Baixa	Nenhuma	<15000	Alto
Ruim	Baixa	Adequada	>35000	Moderado
Boa	Baixa	Nenhuma	>35000	Baixo
Boa	Alta	Adequada	>35000	Baixo
Boa	Alta	Nenhuma	<15000	Alto
Boa	Alta	Nenhuma	>=15000 a <=35000	Moderado
Boa	Alta	Nenhuma	>35000	Baixo
Ruim	Alta	Nenhuma	>=15000 a <=35000	Alto

A partir desse dataset podemos gerar uma outra tabela, a tabela de probabilidade como mostrada abaixo:

	História de crédito			Dívida		Garantia		Renda anual		
Risco de crédito	Boa	Desconhecida	Ruim	Alta	Baixa	Nenhuma	Adequada	<15000	>=15000 a <=35000	>35000
Alto 6/14	5	5	4	7	7	11	3	3	4	7
	1/6	2/6	3/6	4/6	2/6	6/6	0/6	3/6	2/6	1/6
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0/3	2/3	1/3
Baixo 5/14	3/5	2/5	0/5	2/5	3/5	3/5	2/5	0/5	0/5	5/5

Pode-se observar que é pegado a quantidade total de classificação do total de instâncias de classificação, com isso você obtém a probabilidade de cada atributo da tabela e suas instâncias, por exemplo, têm-se um atributo com 2 instâncias SIM e NÃO, essas duas instâncias possuem 5 cada, ou seja, a tabela ficaria para o sim 5/10 e para o NÃO 5/10.

Para avaliar um registro basta fazer os cálculos mostrados na imagem abaixo;

**Registro de teste:**

História: Boa

Dívida: Alta

Garantias = Nenhuma

Renda = >35000

$$P(\text{alto}) = 6/14 * 1/6 * 4/6 * 6/6 * 1/6$$

$$P(\text{alto}) = 0,0079$$

$$P(\text{alto}) = 0,0079/0,0645 * 100 = 12,24\%$$

**Soma:**

$$0,0079 + 0,0052 + 0,0514 = 0,0645$$

$$P(\text{moderado}) = 3/14 * 1/3 * 1/3 * 2/3 * 1/3$$

$$P(\text{moderado}) = 0,0052$$

$$P(\text{moderado}) = 0,0052/0,0645 * 100 = 8,06\%$$

$$P(\text{baixo}) = 5/14 * 3/5 * 2/5 * 3/5 * 5/5$$

$$P(\text{baixo}) = 0,0514$$

$$P(\text{baixo}) = 0,0514/0,0645 * 100 = 79,68\%$$