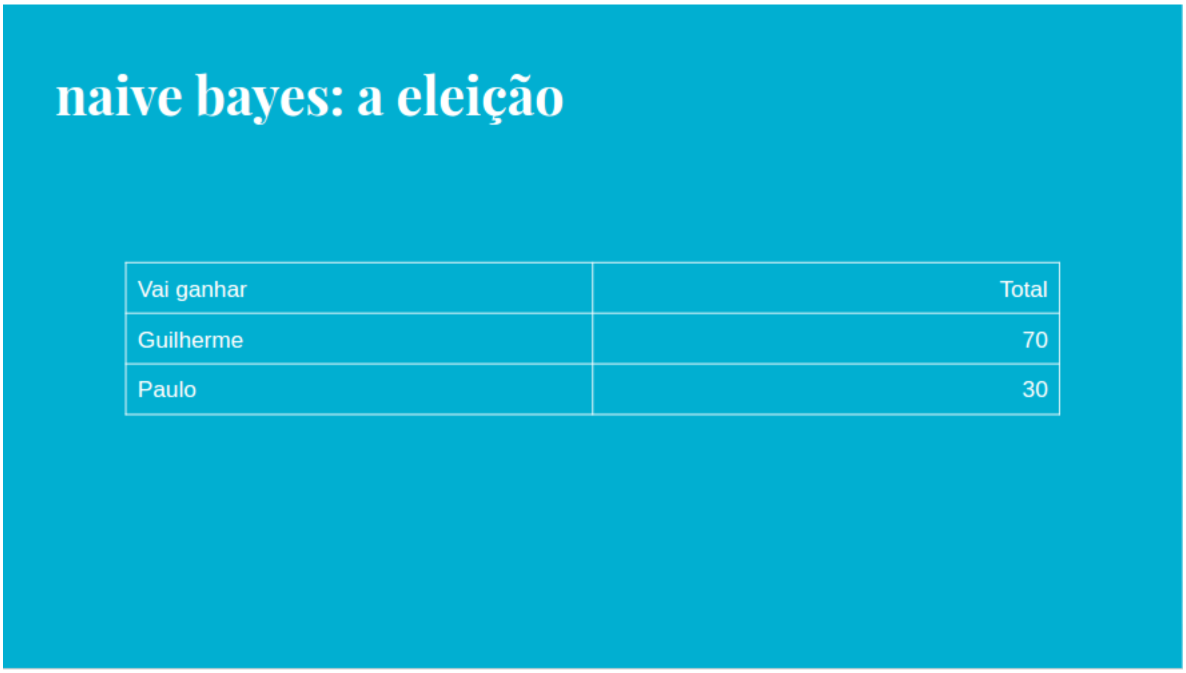
04

Naive bayes e maxium a posteriori por trás dos panos

Naive bayes e maximum a posteriori por trás dos panos

Até agora utilizamos apenas o algoritmo `MultinomialNB` (Naive Bayes). Como será que esse algoritmo funciona por de trás dos panos? Em outras palavras, dado um conjunto de dados que separamos entre treino e teste, como será que ele faz para treinar a partir desses dados? Ou então, como o nosso cérebro faz pra dizer se é um porco ou cachorro? Ou se o e-mail é *spam* ou não *spam*. Mas porque precisamos saber como ele funciona? É justamente para obtermos uma base melhor de como um algoritmo de classificação funciona. Existem diversas formas de classificação que envolve teorias matemáticas, porém, nesse instante, veremos o `MultinomialNB` . Podemos iniciar com um exemplo do dia-a-dia do brasileiro, a eleição. Pegamos um determinado público e perguntamos quem, entre o Guilherme e o Paulo, que iriam ganhar, os resultados foram:



Observe que 70 pessoas acreditam que o Guilherme vai ganhar e 30 pessoas acham que é o Paulo. Nesse exemplo, temos 70 pessoas de 100, que dizem que o Guilherme vai ganhar. Levando essa informação em consideração, se perguntarmos a uma pessoa quem vai ganhar, o que será que ela responderá? A probabilidade para acharem que seja o Guilherme será de 70% e o Paulo 30%. Perceba que a pergunta não tem o objetivo de saber quem vai ganhar e sim, explorar a probabilidade para as duas opções.

Se escolhermos uma pessoa aleatória e perguntarmos quem vai ganhar, utilizando a informação anterior como base, o que você acha que ela vai responder? Lembrando que você só pode falar uma opção, o que você responderia tendo conhecimento dos dados apresentados (treinado com esses dados)? E então? Ela responde Guilherme ou Paulo? Qual foi a regra de decisão que você utilizou para responder entre o Guilherme e o Paulo?

Nesse instante, você está tomando uma decisão de acordo com todo o treinamento que você obteve. Uma das possíveis abordagens para tomada de decisão seria escolher entre o maior deles, ou seja, entre 30% ou 70%, eu escolheria o de 70%, mas porque o de 70%?

naive bayes: a eleição

Vai ganhar	Total
Guilherme	**70%**
Paulo	30%

Porque 70% é maior, portanto, eu, escolheria o maior, nesse caso, o Guilherme. Um outro exemplo seria dizer que 70% das pessoas compram e 30% não compram, novamente eu escolheria o maior, logo, responderia que compram, pois a probabilidade é maior. Note que esse é um exemplo de tomada de decisão. Mas e se pegássemos o de menor probabilidade? Faz sentido em algum momento? A princípio parece ruim, pois se perguntarmos para diversas pessoas, devido a chance de erro ser muito maior, provavelmente não escolherão o Paulo, em outras palavras, se levarmos em consideração todos os chutes, em 70% das vezes as pessoas escolherão o Guilherme, por isso não faz muito sentido utilizarmos a menor probabilidade como parâmetro para tomada de decisão. É importante ressaltar que essa tomada de decisão, que escolhe entre a maior ou a menor probababilidade, é válida apenas para os casos em que analisamos somente a quantidade de acertos. Mas será que não existe alguma outra regra que podemos aplicar?

Perceba que apenas a regra de decisão que vimos até agora é muito rígida, pois estamos utilizando apenas uma informação para decidir quem será ou não votado, ou melhor, quando consideramos apenas o maior dentre as porcentagens, não conseguimos distinguir **70% e 30%** de **99% e 1%** ou **51% e 49%**, ou seja, sempre escolheremos o maior independemente se existe alguma chance ou não para o de menor probabilidade. Repare que restringir a nossa decisão não parece uma boa ideia, ou melhor, podemos melhorar a forma que decidimos qual será o voto escolhido entre o Guilherme e o Paulo.

Uma das possibilidades que podemos abordar seria sortear números entre 1 a 100, então, se o número for menor ou igual a 70, significa que ele faz parte de 70% de chance, ou seja, o voto seria para o Guilherme, caso o contrário, entre 71 a 100, o voto seria para o Paulo que corresponde aos 30%:

naive bayes: a eleição

Vai ganhar	Probabilidade	Na sorte...
Guilherme	70%	1 até 70
Paulo	30%	71 até 100

Mas o que estamos fazendo nesse momento? Estamos criando uma regra que não decide apenas utilizando a maior ou menor probabilidade, em outras palavras, nesse instante, nós temos 70 e 30 números de 100 para decidir se é o Guilherme (70 números) ou o Paulo (30 números). Levando essa nova regra em consideração, se perguntarmos para uma pessoa qualquer em quem ela votará, será que ela vai votar no Paulo ou no Guilherme? Dessa vez, ao invés de escolher o maior para tentar descobrir em quem essa pessoa votará, iremos sortear um número de 1 a 100, pois o número 100 é o número total da probabilidade. Então, se esse número for entre 1 a 70 iremos dizer que essa peossa votou no Guilherme e se for entre 71 a 100 iremos dizer que ela votou no Paulo.

Perceba que agora, não estamos lidando apenas com uma possibilidade como estávamos fazendo antes, dessa vez, existe a possibilidade tanto para o Guilherme ou para o Paulo, independemente se o Guilherme tem mais chance do que o Paulo ou vice-versa. Observe que essa regra funciona de acordo com as probabilidades, pois existe a chance de uma pessoa qualquer escolher ou Guilherme ou Paulo. De acordo com o que vimos até agora, conseguimos definir 3 tipos de regras de decisão:

- **Preferência ao maior**: É válido quando estamos apenas lidando com quantidade.
- **Preferência ao menor**: Não faz sentido, pois o que tiver maiores chances na maioria das vezes será o escolhido.
- **Sortear números aleatórios entre 1 a 100**: É válido quando estamos querendo de fato lidar com as probabilidades.

O nosso treino pode ser baseado em algumas dessas regras de decisões, porém é sempre importante compreender que para cada um dos casos uma dessas regras pode ser mais interessante do que a outra. Um outro exemplo que podemos analisar seria conforme a tabela abaixo:

naive bayes: a eleição

Vai ganhar	Total
Guilherme	77
Paulo	42

Repare que dessa vez, ao invés porcentagens, temos a quantidade de votos do Guilherme e do Paulo. O que podemos fazer quando já temos a quantidade de votos ao invés do percentual? Simples, calculamos o percentual. Para calcularmos o percentual, precisamos saber a quantidade de votos totais, nesse caso, basta somar os votos do Guilherme e do Paulo:

naive bayes: a eleição

Vai ganhar	Total
Guilherme	77
Paulo	42
Total	119

119 é a quantidade total de pessoas que foram entrevistadas e deram os seus votos. Dentre essas 119, 77 votaram no Guilherme e 42 no Paulo. Para calcular a probabilidade, basta pegar a quantidade de votos e dividir pelo total de pessoas, por exemplo, para calcular a probabilidade do Guilherme fazemos `77/119 = 0,647058824` , ou seja, quando arredondamos obtemos 65%. E para o Paulo? A mesma coisa, porém, agora com 42 votos, então, `42/119 = 0,352941176` , ao arredondar esse número obtemos 35%. Então a nossa tabela fica da seguinte maneira:

naive bayes: a eleição

Vai ganhar	Total	Probabilidade	Na sorte...
Guilherme	77	65%	1 e 65
Paulo	42	35%	66 e 100
Total	119		

Se sortearmos um número de 1 a 100 e cair entre 1 e 65, fará parte dos 65%, isto é, o Guilherme será votado, caso o número seja entre 66 a 100, estaremos pegando os 35% do Paulo, logo, o voto vai para o Paulo. Observe que por meio dessa tabela podemos utilizar uma das 3 regras que vimos até agora. Por exemplo, se utilizarmos a regra do **maior**, iremos apenas dizer que o Guilherme será votado, caso optamos pela regra do valor **menor**, falaremos que apenas o Paulo será votado, por fim, podemos escolher a regra da **probabilidade**, em outras palavras, sorteamos um número entre 1 a 100 e então, dependendo do número que for sorteado, ele cairá ou nos 65% do Guilherme (1 a 65) ou nos 35% do Paulo (66 a 100).

Perceba que o comportamento do algoritmo que estamos analisando, é justamente pegar uma quantidade de dados brutos, calcular as probabilidades e utilizar uma regra de decisão. Esses passos são bem similares ao que fazemos no nosso dia-a-dia, por exemplo, vemos diversos animais, e esses animais possuem pernas curtas, são rozinhas, não fazem *auau* entre demais características, então aprendemos que dada as características citadas, a maioria desses animais são porquinhos e a minoria são cachorrinhos, ou seja, calculamos essa probabilidade na nossa cabeça e a partir dela conseguimos concluir se é um porquinho ou um cachorrinho.

Na classificação de e-mails entre *spam* e não *spam* também fazemos algo bem similar, em outras palavras, também utilizamos um dos 3 critérios que vimos até agora, por exemplo, quando vemos os nossos e-mails e nos deparamos com um texto do *subject* que na maioria das vezes era um *spam*, imediatamente, sem utilizar nenhum outro critério, concluímos que se trata de um *spam*, ou então, sabemos que na maioria das vezes aquele *subject* se tratava de um *spam*, mas, por meio de uma escolha aleatória, decidimos abrir esse e-mail, nesse instante estamos entrando na situação que sorteamos o número, e então, verificamos em qual probabilidade ele se encaixa. Um outro exemplo bem comum são as ações da bolsa de valores em que analisamos uma determinada ação e nos perguntamos se ela vai subir ou vai descer, então acontece um evento comum que, na maioria das vezes, a ação subia. Se levarmos em consideração o método do maior valor, investiremos nessa ação, pois acreditamos que ela subirá. Porém, pode também acontecer um evento que em alguns momentos ela subiu ou desceu, então utilizamos o critério da probabilidade, sorteamos um número na nossa cabeça e, a partir número, escolhemos se vamos ou não investir nessa ação.

Observe que até mesmo nós, fazemos passos similares ao do algoritmo de classificação, criando padrões para tomar as nossas decisões, medindo sempre a chance da ocorrência (percentual para que aconteça) e utilizando um método de decisão, seja pelo número maior, menor ou probabilidade. Dentre essas regras de decisão vistas até o momento, a que utilizaremos nos exemplos posteriores, chama-se *maximum a posteriori* que é justamente a regra que, dada as probabilidades, escolhemos o maior valor dentre as possibilidades. Então vamos para o próximo exemplo referente a um site de vendas de imóveis.

No meu site, tenho diversas informações dos clientes que compraram, e dentre elas eu peguei a informação de qual estado que eles são:

naive bayes: a compra

Estado	Total	Probabilidade
Rio de Janeiro	31	
São Paulo	56	
Total		

Podemos ver que 31 dos meus clientes são do Rio de Janeiro, e 56 são de São Paulo. A princípio podemos calcular a probabilidade de uma pessoa ser de São Paulo ou do Rio de Janeiro, porém, o que fazemos com essa informação? A princípio, não temos muita utilidade pra ela... Então o que realmente queremos saber? Na verdade, o que queremos saber é se, dada essas informações, a pessoa vai comprar ou não no meu site. Mesmo a informação que indica qual região os clientes compram mais, ou seja, de São Paulo ou do Rio de Janeiro, tenha alguma relevância, a informação de que o cliente que compra ou não também é de grande importância. Então observe que nesse instante teremos que lidar com duas informações conforme a tabela a abaixo:

naive bayes: a compra, dado que...

Rio de Janeiro	Total	Probabilidade
Comprou	37	
Não comprou	17	
Total		

São Paulo	Total	Probabilidade
Comprou	67	
Não comprou	177	
Total		

Perceba que agora sim as informações já começam a fazer sentido, pois, dado o estado dos meus clientes, temos a informação de quantos compraram ou não. Então agora podemos preencher a nossa tabela com as probabilidades, utilizando o mesmo esquema que fizemos no primeiro exemplo, que é somar a quantidade de clientes que compraram e que não compraram de acordo com o seu estado, então dividimos os que compraram pelo total e obtemos a probabilidade dos clientes que compraram, em seguida, fazemos o mesmo para os que não compraram. O resultado da nossa tabela fica da seguinte forma:

probabilidade condicional...

Rio de Janeiro	Total	Probabilidade
Comprou	37	68,5%
Não comprou	17	31,5%
Total	54	

São Paulo	Total	Probabilidade
Comprou	67	27.5%
Não comprou	177	72.5%
Total	244	

Perceba que agora a nossa tabela possui um aspecto diferente, pois 68,5% dos clientes do Rio de Janeiro compram e 31,5% não compram e, os de São Paulo, 27,5% compram e 72,5% não compram. O que isso significa? Note que agora temos que lidar com duas variáveis que são: **localidade** e **compra**:

- **localidade**: estado que o cliente mora.
- **compra**: se comprou ou não.

Nesse instante precisamos analisar duas informações para poder prevêr se o cliente comprará ou não. Por exemplo, utilizando o método *maximum a posteriori*, um cliente do Rio de Janeiro, compra ou não compra? Como podemos ver 68,5% compram e 31,5% não, em outras palavras, os clientes do Rio de Janeiro irão comprar. Mas e os clientes de São Paulo? 72,5% não compram e 27,5% sim, logo, não irão comprar.

Perceba que essas probabilidades possuem condições, pois de acordo com o estado do cliente, temos uma determinada possibilidade se ele vai comprar ou não, em outras palavras, para cada estado, teremos uma probabilidade de compra dele, por exemplo, se o cliente for do Rio de Janeiro ele tem mais chance de comprar, porém, se for de São Paulo ele tem menos chance de comprar. Esse é o tipo de probabilidade quem chamamos de **probabilidade condicional**.

Levando em consideração esse tipo de probabilidade, se escolhêssemos uma pessoa qualquer para verificar se ela comprou ou não, qual seria a nossa atitude? Primeiramente, perguntaríamos o estado dela, então, se fosse do Rio de Janeiro, responderíamos que ela compraria, porém, se fosse de São Paulo, falaríamos que ele não iria comprar, baseando-se no método do *maximum a posteriori*. Mas e se usássemos a regra da probabilidade? Como faríamos? Da mesma forma! A única diferença é que ao invés de escolher o maior, iríamos sortear um número de 1 a 100 e então verificaríamos em qual probabilidade ele se encaixa, por exemplo, se fosse o Rio de Janeiro, de 1 a 68,5 responderíamos que compraria, porém, de 68,6 a 100 diríamos que não compraria.

Observe que para esses tipos de dados (com probabilidades condicionais), podemos analisar diversas perguntas, como por exemplo:

- Dado que ele é do Rio de Janeiro, ele vai comprar?

Qual é a nossa atitude dessa vez? Primeiro verificamos em qual tabela ele pertence, nesse caso, Rio de Janeiro, logo, ele tem 68,5% de chance de comprar, então podemos responder que sim. A próxima pergunta seria:

- Dado que ele é do Rio de Janeiro, qual a probabilidade de ele comprar?

E para essa situação? Fazemos a mesma coisa, porém respondemos a chance, ou seja, probabilidade desse cliente comprar que é 68,5% . Perceba que na matemática podemos representar essa mesma pergunta da seguinte forma:

```
P(Comprar | Rio de Janeiro)
```

Repare que a matemática só traduz, ou seja, P significa probabilidade e `(comprar | Rio de Janeiro)` nos diz que, probabilidade de comprar, dado que ele é do Rio de Janeiro. Então podemos concluir que essa representação matemática é a tradução da nossa pergunta. E se fosse de São Paulo? Como ficaria? Vejamos a próxima pergunta:

- Dado que ele é de São Paulo, qual a probabilidade de ele comprar?

Da mesma forma que fizemos para o Rio de Janeiro, representamos matematicamente da seguinte fórmula:

```
P(Comprar | São Paulo)
```

Até o momento, tivemos apenas 1 única variável condicional que indica qual é o estado. Então, de acordo com o estado olhamos a probabilidade de comprar ou não. Representamos essa probabilidade condicional conforme a tabela abaixo:

Com apenas uma váriavel condicional utilizamos essas duas tabelas para nos auxiliar. Porém, e se forem duas variáveis condicionais? Por exemplo, sabemos qual é o estado do cliente e, além do estado, também temos a informação sobre sua renda, em outras palavras, sabemos também se a renda do cliente é maior ou menor do que 5000. Esse tipo de informação influencia se o cliente vai ou não comprar, pois de acordo com o mercado, os clientes com maiores condições possuem uma chance maior de comprar do que os que possuem menores condições entre outros estudos sobre o mercado. Tendo o conhecimento das informações apresentadas anteriormente, nos deparamos com a seguinte situação:

- Temos duas variáveis condicionais, o estado e a faixa salarial do cliente, como podemos fazer a nossa analise?

Observe que agora temos duas variáveis condicionais, ou seja, precisamos analisar duas variáveis distintas, mas como podemos resolver isso? Da mesma forma que fizemos para a variável que representa o estado, iremos criar uma tabela para a variável que representa a faixa de salário:

Então repara que agora temos 4 tabelas, dentre elas, duas representam os estados em que os clientes moram e as outras duas, representam a faixa salarial dos nossos clientes. O que precisamos fazer agora? Preencher a tabela da faixa salarial:

Veja que, 80% dos clientes com salário acima de 5000 compram e 20% não, porém, os clientes com salários abaixo de 5000 18% compram e 82% não. Por meio dessa tabela, podemos concluir que o perfil de uma pessoa que ganha mais que 5000 é totalmente diferente de uma pessoa que ganha menos de 5000. Pode ser que os imóveis que estamos vendendo sejam mais caros, ou então, nesse instante, os apartamentos que eram mais caros estão com valores mais acessíveis... Podem existir deversos pontos para explicar o motivo para alguns clientes, com um determinado perfil, comprar mais do que outros. De fato não sabemos dizer os reais motivos para uns comprarem mais do que outros, porém, sabemos que os clientes com faixa de salário maior que 5000, possuem maiores chances de comprar do que os clientes com faixa abaixo de 5000, como também sabemos dizer que, dentre os clientes do Rio de Janeiro ou de São Paulo, os clientes com maior chance de comprar são os do Rio de Janeiro.

Repare que isolamos as informações de estado e faixa de salário, ou seja, elas são independentes, mas, agora que temos as informações, o que devemos fazer? Lembra que treinávamos o nosso algoritmo? Então agora precisamos treinar também, com base nessas informações. E como treinamos a partir dessas informações? Simplesmente passamos por cada uma das tabelas, verificamos a sua variável condicional e seu valor, e então, coletamos suas respectivas probabilidades de comprar ou não. Vejamos o resultado desse treino:

- 1º tabela: `estado:` Rio de Janeiro.
  - 68,5% compram e 31,5% não compram.
- 2º tabela: `estado:` São Paulo.
  - 27,5% compram e 72,5% não compram.
- 3º tabela: `salário:` $ > 5000.
  - 80% compram e 20% não compram.
- 4º tabela: `salário:` $ < 5000.
  - 18% compram e 82% não compram.

Agora que treinamos, o que precisamos fazer? Testar! E como fazemos o teste com base no nosso treino? Da mesma forma que fizemos antes, em outras palavras, iremos calcular a chance e utilizar uma regra de decisão, porém, a única diferença é que pra cada variável condicional que tivermos, precisaremos criar duas tabelas, isto é, quando tinhamos apenas a variável para o estado, criamos apenas duas tabelas, agora que temos a variável estado e faixa salarial, criamos quatro, se fossem três variáveis, seis tabelas e assim por diante. Começaremos com o seguinte exemplo:

- Dado um cliente que é de São Paulo com salário menor do que 5000. Qual é a probabilidade de ele comprar?

Lembra que podemos representar essa pergunta com uma fórmula matemática? Porém tinhamos feito apenas para uma variável condicional. Como ficaria a mesma fórmula para duas variáveis, ou seja, as variáveis estado e faixa de slário? Vejamos:

```
P(Comprar | São Paulo, <= 5000)
```

Perceba que por meio da fórmula, temos os valores de ambas as variáveis, nesse caso, estado e salário, e, além disso, já sabemos quais são os valores, ou seja, São Paulo e menor do que 5000 respectivamente. Mas como calcularemos essa probabilidade que envolve tanto um estado e faixa de salário a partir dessa fórmula?

Podemos pensar da seguinte maneira: "Já que as **variáveis são independentes**, podemos calculá-las individualmente"! Isso significa que, devido ao fato de ambas as variáveis não depender uma da outra, podemos quebrar a fórmula matemática mais complexa (`P(Comprar | São Paulo, <= 5000)`), em duas fórmulas mais simples:

```
P(Comprar | São Paulo, <= 5000) = P(Comprar | São Paulo) * P(Comprar | <= 5000)
```

Observe que criamos duas fórmulas distintas, em outras palavras, uma não depende da outra. A partir de agora, estamos calculando a probabilidade separadamente! Então começaremos pela fórmula `P(Comprar | São Paulo)` :

- `P(Comprar | São Paulo)` : Qual é a probabilidade de compra para os clientes de São Paulo?
    - 27,5% compram.
    - 72,5% não compram.

Agora para a fórmula `P(Comprar | <= 5000)` :

- `P(Comprar | <= 5000)` : Qual é a probabildade de compra para os clientes que ganham menos de 5000?
    - 18% compram.
    - 82% não compram.

Por fim, multiplicamos ambas as probabilidades e obtemos a nossa probabilidade final, por exemplo:

- Qual a probabilidade de um cliente de São Paulo comprar?
    - 27,5%.
- Qual a probabilidade de um cliente que ganha menos de 5000 comprar?
    - 18%.

Temos as porcentagens 27,5% e 18%, então fazemos:

```
27,5% * 18%
```

Então obtemos o seguinte resultado:

```
0,275 * 0,18 = 0,0495 -> 100 * 0.0495 = 4,95 = 5%
```

Observe que a probabilidade de um cliente comprar dado que ele seja de São Paulo e tenha um salário menor que 5000 foi de 4,95%, porém, arredondamos para 5% por ser um valor muito próximo, ou seja, o nosso teste resultou em 5%. Agora que calculamos, precisamos utilizar uma regra de decisão. Qual regra seria interessante? A de maior valor? Vejamos:

- 5% compram.

- 95% não compram.

Repare que se utilizarmos a regra do maior esse cliente não irá comprar, pois existe 95% de chance dele não comprar. E se tentarmos a regra de probabilidade? Teremos que sortear um número de 1 a 100, se o resultado for de 1 a 5 o cliente vai comprar, caso o resultado for de 6 a 100, ele não vai comprar. Como podemos ver, o resultado final dependerá bastante da regra de decisão que escolhermos. Vejamos um outro exemplo:

```
P(Comprar | Rio de Janeiro, >5000)
```

Um cliente que é do Rio de Janeiro e tem um salário acima de 5000. Dado que esse cliente é do Rio de Janeiro e ganha mais de 5000, qual é a probabilidade dele ganhar? Faremos da mesma forma como anteriormente, ou seja, calcularemos cada probabilidade de acordo com sua variável:

- Qual a probabilidade dele comprar dado que ele é do Rio de Janeiro?
  - 68,5%
- Qual a probabilidade dele comprar dado que ele ganha mais de 5000?
  - 80%

Então multiplicamos as probabilidades:

```
0,685 * 0,80 = 0,548 -> 100 * 0,548 = 54,80%
```

Veja que um cliente que é do Rio de Janeiro e ganha acima de 5000 contém 54,8% de chance de comprar. Um resultado bem diferente comparado ao cliente que era de São Paulo e ganhava menos de 5000 que continha apenas 5% de chance de comprar.

Resumindo

Perceba que podemos calcular a probabilidade de um evento, como por exemplo, um cliente que é de um estado `X` e ganha um salário ou maior ou menor que `Y`. Fizemos isso multiplicando todas as variáveis existentes para obter o resultado final, pois assumidos que ambas as variáveis são independentes. Esse é o processo de treino que precisamos realizar, ou seja, calcular as probabilidades condicionais independentemente se é uma ou duas ou três variáveis condicionais. Para cada variável criamos duas tabelas, isto é, se for apenas uma variável, então duas tabelas, se forem duas variáveis, então quatro tabelas, se forem `X` variáveis, então `X * 2` tabelas.

Então utilizamos um critério de avaliação, podemos utilizar tanto o `maximum a posteriori` quanto a probabilidade, porém, precisamos sempre nos atentar a qual critério escolher, pois, por exemplo, se utilizarmos o `maximum a posteriori` e as probabilidades de compra forem sempre maiores, ele sempre vai dizer que o cliente vai comprar, como também, caso o contrário, ou seja, todas a probabilidades de compra forem menor, ele sempre irá dizer que o cliente não vai comprar. Percebe o quão problemático ele pode ser? É exatamente por esse motivo que, na maioria dos casos, utilizar o critério da probabilidade (que sorteia os números de 1 a 100) acaba sendo mais adequado para treinar o nosso algoritmo. Mas e o nosso teste? Como funciona? É simples, multiplicamos todas as variáveis condicionais que foram calculadas e aplicamos uma regra de decisão, então, tomamos uma decisão, em outras palavras, classificamos se o elemento é da categoria A ou B, se é *spam* ou não, se ele vai comprar ou não.

Repare que utilizamos o algoritmo *Multinomial naive bayes* que se caracteriza por ser um treino extremamente rápido, pois é só calcular essas tabelas que vimos no exemplo. Além disso, ele precisa apenas realizar uma somatória, em outras palavras, precisa somar, dado um determinado elemento, quantos deram uns e quantos deram zeros. Esse tipo de algoritmo é chamado de [linear (https://pt.wikipedia.org/wiki/Programa%C3%A7%C3%A3o_linear)](https://pt.wikipedia.org/wiki/Programa%C3%A7%C3%A3o_linear), portanto, podemos concluir que ele é um algoritmo eficiente, pois ele faz operações de acordo com a quantidade de elementos, por exemplo, se são 100 elementos ele faz apenas as 100 operações, se forem 1000 elementos, 1000 operações, ou seja, ele tem um desempenho muito bom comparado com outros algoritmos existentes. Vimos também que ele é simples de implementar, pois ele realiza algumas operações matemáticas que são equivalentes, teoricamente, com as tabelas que analisamos.

Além disso, ele é bastante utilizado para classificar texto, como foi o caso do *spam*, por exemplo, verificar as palavras mais frequentes e, com base nessas informações, ele tenta distinguir um e-mail normal de um *spam*. Perceba o *naive bayes* não é uma implementação de um outro mundo, pois ele está relacionado com determinados passos e rotinas que temos uma pequena noção, em outras palavras, temos a capacidade de analisar situações, como por exemplo, dado que 70% das pessoas vão recomendar o produto e 30% não, e então, perguntarmos para uma pessoa qualquer se ela recomenda ou não o produto, provavelmente ela dirá que sim. Podemos observar que essa é uma sensação que todos nós, seres humanos, compartilhamos, como por exemplo, se, com base na informação: "A maior parte dos brasileiros gostam de futebol". Levando em consideração essa afirmação, se perguntarmos para um brasileiro qualquer se ele gosta de futebol, o que você responderia? Provavelmente responderia sim, pois é a maior parte das pessoas, é bem provável que também terão pessoas que não gostam, mas não importa, pois estamos dando preferência a quantidade maior, ou seja, utilizando a regra do maior, ou então, poderíamos utilizar uma outra regra de decisão.

Perceba que o senso de eventos mais comuns, em outras palavras, com maior frequência de ocorrência, é uma opção válida para uma regra de decisão e é muito parecido com as regras que estamos utilizando, ou seja, o que tiver maior frequência provavelmente acontecerá denovo. Esse é justamente o comportamento do *naive bayes*, de acordo com a probabilidade que aconteceu no passado, eu irei conseguir classificar o que acontecerá agora.