***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:

Texto

Descrição gerada automaticamente com confiança média

Texto preto sobre fundo branco

Descrição gerada automaticamente

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***

Tabela

Descrição gerada automaticamente

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 somente aplicando em casos onde a história de crédito for boa***:

* 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:

***Obs: O exemplo acima serve apenas de ilustração, você ainda precisará fazer o modelo para as colunas desconhecida, e ruim da história de crédito***

***Figura 1 – Teorema de Bayes Exemplo***

Tabela

Descrição gerada automaticamente

Após acabar a análise probabilística de uma coluna é necessário trazer as outras variáveis para o modelo, conforme os exemplos abaixo.

***Figura 2 – Teorema de Bayes Atributos Dívida X Risco***

***Tabela

Descrição gerada automaticamente***

***Figura 3 – Teorema de Bayes Atributos Garantia X Risco***

***Tabela

Descrição gerada automaticamente***

***Figura 4 – Teorema de Bayes Atributos Renda Anual X Risco***

***Tabela

Descrição gerada automaticamente***

Com as imagens acima, concluímos a tabela probabilística de Naive Bayes, mas um ponto que precisamos sempre lembrar no momento de validação dessa análise, é se os valores estão coerentes

Se temos 6 registros classificados como risco alto. Precisamos ter esses registros distribuídos pelas variáveis

Exemplo: a soma do risco por história de crédito alta

***É 1/6 quando o cliente tem história de crédito boa***

***2/6 quando o cliente tem história de crédito moderado***

***3/6 Quando o cliente tem história de crédito ruim***

A soma desses valores dá os 6 registros altos que temos na coluna de classificação, logo é correto. Mas precisamos nos certificar que as somas das outras colunas também façam sentido com os valores existentes.

***Cálculo do Naive Bayes***

Para entendermos sobre a probabilidade de risco do cliente precisamos fazer um insert dos dados do cliente.

Exemplo de massa de dados simples:

* O cliente tem história de crédito Boa
* O cliente tem Dívida Alta
* Garantia Nenhuma
* Renda >= 35K

Tabela

Descrição gerada automaticamente

Uma vez com essas informações, iremos selecionar apenas os casos em que notamos os dados inseridos do teste de mesa

Exemplo:

Uma imagem contendo Linha do tempo

Descrição gerada automaticamente

E a partir dessa seleção iremos multiplicar por cada um dos atributos. Pegando o total de vezes que temos um registro alto, por exemplo, e multiplicando pelos atributos do cliente.

* Ou Seja, como ele tem história de crédito boa, seria o equivalente a 1/6
* Como ele tem divida alta, seria o equivalente a 4/6
* Como ele não tem nenhuma garantia, seria o equivalente a 6/6
* Como ele tem uma renda anual > 35K seria o equivalente a 1/6

Exemplo do cálculo:

Para probabilidade Alta:

P(ALTO) = 6/14 \* 1/6 \* 4/6 \* 6/6 \* 1/6

No entanto, importante lembrar que precisamos fazer para todas as classificações, como o moderado e baixo também.

***Figura Cálculo Naive Bayes para todos os riscos***

Linha do tempo

Descrição gerada automaticamente

Na figura acima, deixo os cálculos por probabilidade do cliente com certos atributos, assim como o cálculo da probabilidade de risco do cliente em porcentagem

Notamos que para

P(BAIXO) = 79,68%

P(Moderado) = 8,06%

P(Baixo) = 12,24%