Algoritmo Naive Bayes é baseado em probabilística, no teorema Bayes.

O objetivo do algoritmo de Naive Bayes é gerar tabela de probabilidade a partir da tabela (dataset/histórico) original.

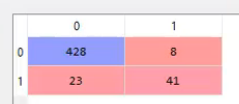
Correção laplaciana = forma de tratar registros que tem valores zerados, adicioionando um novo registro.

Probabilidade apriori = valor de probabilidade conhecido

Probabilidade posteori = valor de probabilidade resultante.

Você usa probabilidade apriori para determinar posteori

Naive Bayes não aceita atributo categóricos. Logo precisca converte-los.



A matriz de confusão acima expressa relação das classes 0 (não pagou) e 1 (pagou), no qual a intersecção entre essas classes da o valor de registro que o algoritmo classificou incorretamente ou corretamente. No caso, ficou da seguinte forma:

* intersecção de 0 com 0 significa que 428 foram classificados corretamentes
* intersecção de 0 com 1 significa que 8 foram classificados como 1 porém deveria ter sido classificado como 0.
* intersecção 1 com 0 significa que 23 foram classificados como 0 porém deveriam ter sido classificados como 1.
* intersecção 1 com 1 significa que 48 foram classificados corretamentes.

Nem sempre seguindo a literatura, fazendo todos possíveis pre processamento indicados que se obterá um algoritmo com uma maior acurácia, ás vezes é melhor não fazer determinado pre-processamento que esse não fazer aumentará acuracia do algoritmo. Por exemplo, escalonamento em variaveis dumb para algoritmo naive bayes não é uma boa, na verdade faz perder acuracia.