Naive Bayes:

Bayes Ağı, bir dizi değişken arasındaki ilişkilerin olasılıklarını öğrenerek çıkarım yapan bir makine öğrenmesi algoritmasıdır. Düşünce ya da oluşan kanı veya yönlendirilmiş olasılıklı bir modeldir. Örneğin, bir Bayes ağı, hastalıklar ve semptomlar arasındaki olasılık ilişkilerini temsil edebilir. Belirtiler verildiğinde, ağ çeşitli hastalıkların varlığının olasılıklarını hesaplamak için kullanılabilir.

Naive Bayes sınıflandırıcıları, Bayes Teoremine dayanan bir sınıflandırma algoritmaları koleksiyonudur. Bu tek bir algoritma değil, hepsinin ortak bir prensibi paylaştığı bir algoritma ailesidir; yani sınıflandırılan her özellik çifti birbirinden bağımsızdır.

En basit ve etkili sınıflandırma algoritmalarından biri olan Naïve Bayes sınıflandırıcı, hızlı tahmin yetenekleriyle makine öğrenimi modellerinin hızlı bir şekilde geliştirilmesine yardımcı olur. Sınıflandırma problemlerinde Naïve Bayes algoritması kullanılmaktadır. Metin sınıflandırmada oldukça kullanılır. Metin sınıflandırma görevlerinde veriler yüksek boyut içerir (çünkü her kelime verideki bir özelliği temsil eder). Spam filtrelemede, duyarlılık tespitinde, derecelendirme sınıflandırmasında vb. kullanılır. Naif Bayes kullanmanın avantajı hızıdır. Hızlıdır ve yüksek veri boyutuyla tahmin yapmak kolaydır. Bu model, bir örneğin belirli bir özellik değeri kümesine sahip bir sınıfa ait olma olasılığını tahmin eder. Olasılığa dayalı bir sınıflandırıcıdır. Çünkü modeldeki bir özelliğin başka bir özelliğin varlığından bağımsız olduğunu varsaymaktadır. Başka bir deyişle, her bir özellik birbiriyle hiçbir ilişkisi olmadan tahminlere katkıda bulunmaktadır. Gerçek dünyada bu koşul nadiren karşılanır. Eğitim ve tahmin için algoritmada Bayes teoremini kullanır

Örnek:

Gün	Görünüş	Sıcaklık	Nem	Rüzgar	Play
1	Güneşli	Sıcak	Yüksek	Zayıf	Hayır
2	Güneşli	Sıcak	Yüksek	Kuvvetli	Hayır
3	Bulutlu	Sıcak	Yüksek	Zayıf	Evet
4	Yağmurlu	Hafif	Yüksek	Zayıf	Evet
5	Yağmurlu	Soğuk	Normal	Zayıf	Evet
6	Yağmurlu	Soğuk	Normal	Kuvvetli	Hayır
7	Bulutlu	Soğuk	Normal	Kuvvetli	Evet
8	Güneşli	Hafif	Yüksek	Zayıf	hayır
9	Güneşli	Soğuk	Normal	Zayıf	Evet
10	Yağmurlu	Hafif	Normal	Zayıf	Evet
11	Güneşli	Hafif	Normal	Kuvvetli	Evet
12	Bulutlu	Hafif	Yüksek	Kuvvetli	Evet
13	Bulutlu	Sıcak	Normal	Zayıf	Evet
14	Yağmurlu	Hafif	Yüksek	Kuvvetli	hayır
	Güneşli	**Soğuk**	**Yüksek**	**Kuvvetli**	?

Burada ;

Görünüş, sıcaklık

Nem ve rüzgar özellik matrisidir.

Yanıt vektörü ise

EVET veya HAYIR’dır.

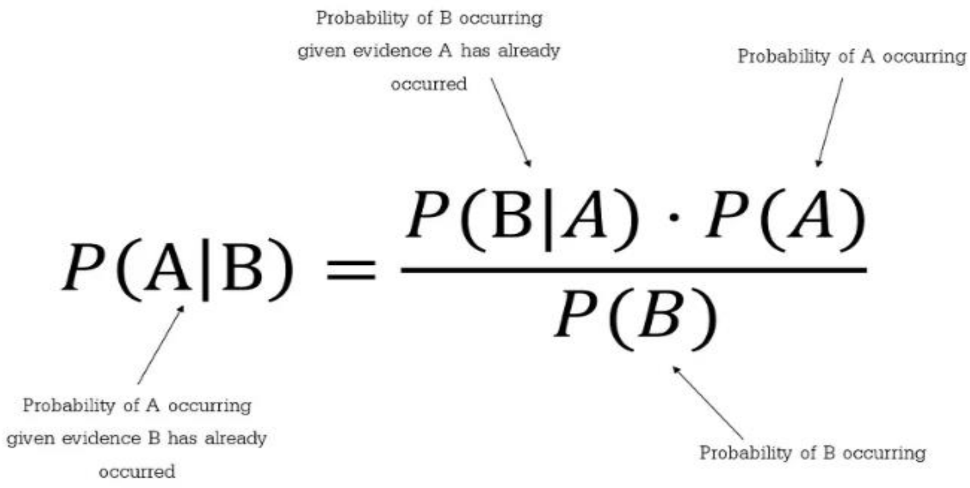
Veri seti **özellik matrisi** ve **yanıt vektörü** olmak üzere iki kısma ayrılmıştır:

Özellik matrisi, her vektörün bağımlı özelliklerin değerinden oluştuğu veri kümesinin tüm vektörlerini (satırlarını) içerir. Yukarıdaki veri setinde özellikler 'Görünüm', 'Sıcaklık', 'Nem' ve 'Rüzgar'dır.

Yanıt vektörü, özellik matrisinin her satırı için sınıf değişkeninin (tahmin veya çıktı) değerini içerir. Yukarıdaki veri setinde sınıf değişkeninin adı 'Golf oyna'dır.

Bayes teoremi:

Bayes Teoremi, daha önce meydana gelmiş başka bir olayın olasılığı göz önüne alındığında, bir olayın meydana gelme olasılığını bulur. Bayes teoremi matematiksel olarak aşağıdaki denklemle ifade edilir:



Bayes Theorem

burada A ve B olaylardır ve P(B) ≠ 0

Temel olarak, B olayının doğru olması durumunda A olayının olasılığını bulmaya çalışıyoruz. B olayı aynı zamanda delil olarak da adlandırılır.

P(A), A'nın önselidir (önsel olasılık, yani olayın kanıt görülmeden önceki olasılığı). Kanıt, bilinmeyen bir örneğin öznitelik değeridir (burada B olayıdır).

P(B) Marjinal Olasılıktır: Kanıt Olasılığı.P(A|B), B'nin posteriori olasılığıdır, yani delil görüldükten sonra olayın olasılığıdır.

P(B|A) Olasılık olasılığıdır, yani kanıtlara dayalı olarak bir hipotezin gerçekleşme olasılığı.

Şimdi Yukarıdaki Örneği İnceleyelim:

Hayır=5

P(Evet)=9/14
P(Hayır)=5/14

Havanın görünüş durumuna göre,
Havanın görünüşü güneşli olduğunda 2 gün oyun var. Hava güneşli olduğunda oyun oynama olasılığı, P(H_Güneşli | Evet)=2/9
Havanın görünüşü güneşli olduğunda 3 gün oyun yok. P(H_Güneşli | Hayır)=3/5
Toplam güneşli gün sayısı=2+3=5

Havanın görünüşü bulutlu olduğunda 4 gün oyun var. P(H_Bulutlu | Evet)=4/9
Havanın görünüşü bulutlu olduğunda 0 gün oyun yok. P(H_Bulutlu | Hayır)=0/5
Toplam bulutlu gün sayısı=4+0=4

Havanın görünüşü yağmurlu olduğunda 3 gün oyun var. P(H_Yağmurlu | Evet)=3/9
Havanın görünüşü yağmurlu olduğunda 2 gün oyun yok. P(H_Yağmurlu | Hayır)=2/5
Toplam yağmurlu gün sayısı=2+3=5
Toplam gün sayısı=5+4+5=14

Sıcaklık durumuna göre,
Sıcaklığın sıcak olduğunda 2 gün oyun var. P(S_Sıcak | Evet)=2/9
Sıcaklığın sıcak olduğunda 2 gün oyun yok. P(S_Sıcak | Hayır)=2/5
Toplam sıcak gün sayısı=2+2=4

Sıcaklığın hafif olduğunda 4 gün oyun var. P(S_Hafif | Evet)=4/9
Sıcaklığın hafif olduğunda 2 gün oyun yok. P(S_Hafif | Hayır)=2/5
Toplam hafif gün sayısı=4+2=6

Sıcaklığın soğuk olduğunda 3 gün oyun var. P(S_Soğuk | Evet)=3/9
Sıcaklığın soğuk olduğunda 1 gün oyun yok. P(S_Soğuk | Hayır)=1/5
Toplam soğuk gün sayısı=3+1=4
Toplam gün sayısı=4+6+4=14

Rüzgar durumuna göre,
Rüzgar zayıf olduğunda 6 gün oyun var. P(R_Zayıf | Evet)=6/9=2/3
Rüzgar zayıf olduğunda 2 gün oyun yok. P(R_Zayıf | Hayır)=2/5
Toplam rüzgar zayıf gün sayısı=6+2=8

Rüzgar Kuvvetli olduğunda 3 gün oyun var. P(R_Kuvvetli | Evet)=3/9=1/9
Rüzgar Kuvvetli olduğunda 3 gün oyun yok. P(R_Kuvvetli | Hayır)=3/5

Toplam rüzgar kuvvetli gün sayısı=3+3=6
Toplam gün sayısı=8+6=14

Nem durumuna göre,
Nem yüksek olduğunda 3 gün oyun var. P(N_Yüksek | Evet)=3/9
Nem yüksek olduğunda 4 gün oyun yok. P(N_Yüksek | Hayır)=4/5
Toplam nem yüksek gün sayısı=3+4=7

Nem normal olduğunda 6 gün oyun var. P(N_Normal | Evet)=6/9
Nem normal olduğunda 1 gün oyun yok. P(N_Normal | Hayır)=1/5
Toplam nem zayıf gün sayısı=6+1=7
Toplam gün sayısı=7+7=14

X={Güneşli, Soğuk, Yüksek, Kuvvetli} ise

P(X | Evet) = P(Evet) * P(H_Güneşli | Evet) * P(S_Soğuk | Evet) * P(R_Kuvvetli | Evet) * P(N_Yüksek | Evet)

P(X | Evet) = (9/14)*(2/9)*(1/9)*(3/9)=(1/7)*(1/27)=1/189=0.053

P(X | Hayır) = P(Hayır) * P(H_Güneşli | Hayır) * P(S_Soğuk | Hayır) * P(R_Kuvvetli | Hayır) * P(N_Yüksek | Hayır)

P(X | Hayır) = (5/14)* (3/5)*(1/5)*(3/5)*(4/5)=(3/14)(12/125)=18/875=0.0206

3 farklı Naive Bayes modeli vardır;

1. Gaussian Naive Bayes
2. Multinomial Naive Bayes
3. Bernoulli Naive Bayes