

# Naive Bayes

- Olasılık:
- Subjektif Olasılık



Sözelimi sabah evden çıkıyorsunuz, gökyüzüne bakarak, havayı bulutlu gördüyseniz yağmur yağma olasılığını yüksek buluyor ve şemsiyenizi yanınıza alıyorsunuz. Yağmur yağarsa, olasılığı doğru tahmin etmiş oluyor ve ıslanmıyorsunuz; ancak eğer yağmur yağmazsa olasılık tahmininiz başarısız oluyor ve bütün gün şemsiyenizi boşa taşımış oluyorsunuz.

Benzer şekilde, öğrencisiniz bir sınava gireceksiniz, çok fazla vaktiniz yok, bütün ders notlarına bakmanız mümkün değil. Öğretim üyesinin derste daha fazla üzerinde durduğu konuları potansiyel sınav sorusu olarak görüyor ve bu konulardan soru çıkma olasılığını yüksek olarak değerlendiriyor ve sadece bu konulara çalışıyorsunuz. Olasılık tahmininiz başarılıysa, sınavınız iyi geçiyor ya da tersi.

Tuttuğunuz futbol takımının maçı var, takımın daha önceki maçlarda performansı muhteşem. Rakip takımla yapılmış daha önceki maçlarda hep üstünlük sağlanmış. Bu maçı da kazanma olasılığı sizce çok yüksek. Bunlar hep içinde bulunulan mevcut durum ve geçmişin bir değerlendirmesi sonucunda sizde oluşan beklenti. Maçın sonucu ise ya beklediğiniz gibi galibiyetle sonuçlanır; ya da tersi.

# Olasılık

- Objektif olasılık, sübjektif olasılık gibi kişiden kişiye değişmez. Olasılık kuramı gereği, belirli bir olay için, belirli koşullar altında tek bir olasılık değeri vardır ve bu değer herkes için aynıdır.
- Sözelimi, bir tarafı yazı diğer tarafı tura olan hilesiz bir parayı havaya atmanız halinde, yazı gelme olasılığının  $\frac{1}{2}$  olduğunu düşünürsünüz. Bu durum sizin için de, dünya üzerinde yaşayan tüm insanlar için de aynıdır ve  $\frac{1}{2}$ 'dir.

$$\text{Olasılık} = \frac{\text{Elverişli sonuç sayısı}}{\text{Toplam sonuç sayısı}}$$

$$P = \frac{m}{n}$$



# Olasılık

Söz gelimi, bir paranın atılması deneyinde evrensel küme;

$$E = \{Yazı, Tura\}$$

Bir zarın atılması deneyinde ise evrensel küme;

$$E = \{1,2,3,4,5,6\}$$

olmaktadır.

İki paranın birlikte atılması deneyinde evrensel küme;

$$E = \{YazıYazı, YazıTura, TuraYazı, TuraTura\}$$

Zar attık 3 gelme olasılığı?  
Zar attık 6 gelme olasılığı?  
Zar attık çift gelme olasılığı?

$$Olasılık = \frac{\text{Elverişli sonuç sayısı}}{\text{Mümkün sonuç sayısı}}$$

$$P = \frac{m}{n}$$

# Olasılık

Söz gelimi, bir paranın atılması deneyinde evrensel küme;

$$E = \{Yazı, Tura\}$$

Bir zarın atılması deneyinde ise evrensel küme;

$$E = \{1,2,3,4,5,6\}$$

olmaktadır.

İki paranın birlikte atılması deneyinde evrensel küme;

$$E = \{YazıYazı, YazıTura, TuraYazı, TuraTura\}$$

Zar attık 3 veya 6 gelme olasılığı?

Zar attık 3 gelme olasılığı?  
Zar attık 6 gelme olasılığı?  
Zar attık çift gelme olasılığı?

$$Olasılık = \frac{\text{Elverişli sonuç sayısı}}{\text{Mümkün sonuç sayısı}}$$

$$P = \frac{m}{n}$$

# Olasılık

Söz gelimi, bir paranın atılması deneyinde evrensel küme;

$$E = \{Yazı, Tura\}$$

Bir zarın atılması deneyinde ise evrensel küme;

$$E : P(A \cup B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

olmaktadır.

İki paranın birlikte atılması deneyinde evrensel küme;

$$E = \{YazıYazı, YazıTura, TuraYazı, TuraTura\}$$

Zar attık 3 veya 6 gelme olasılığı?

$$P(A \cup B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Zar attık 3 gelme olasılığı?

$$P(A) = \frac{1}{6}$$

Zar attık 6 gelme olasılığı?

$$P(B) = \frac{1}{6}$$

Zar attık çift gelme olasılığı?

$$\text{Olasılık} = \frac{\text{Elverişli sonuç sayısı}}{\text{Mümkün sonuç sayısı}}$$

$$P = \frac{m}{n}$$

# Koşullu Olasılık

- A ve B birbiri ile bağdaşan ve gerçekleşme olasılığı bulunan iki basit olay olsun. A ve B olaylarının kesişim kümesinin boş küme olmaması yani ortak elemanı bulunması halinde, A olayının kesinlikle meydana gelmesi durumunda B olayının da meydana gelmiş olma olasılığına koşullu olasılık denir.  $P(A) \neq 0$  ve  $P(B) \neq 0$  ve  $P(A \cap B) \neq \emptyset$

A olayının koşul olması durumunda B olayının gerçekleşme olasılığı:  $P(B/A)$ ,

koşul

B olayının koşul olması durumunda A olayının gerçekleşme olasılığı ise:  $P(A/B)$

A olayının gerçekleşmesi halinde B olayının da gerçekleşmiş olma olasılığı,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

B olayının gerçekleşmiş olması durumunda A olayının gerçekleşmiş olma olasılığı,

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

# Koşullu Olasılık

Bir zarın atılması deneyinde, A olayı zarın üç veya üçten küçük sayı gelmesi, B olayı ise, zarın altıdan küçük çift sayı gelmesi olarak tanımlanıyor. Buna göre, A koşulken B olayının olasılığını,  $[P(B/A)]$  ve B olayı koşulken A olayının olasılığını,  $[P(A/B)]$  hesaplayınız.

Öncelikle A ve B olaylarının elverişli sonuçlarını göstererek olasılıklarını hesaplayalım:

$$A = \{1,2,3\} \quad \text{ve} \quad P(A) = \frac{3}{6}$$

$$B = \{2,4\} \quad \text{ve} \quad P(B) = \frac{2}{6}$$

$$A \cap B = \{2\} \quad \text{ve} \quad P(A \cap B) = \frac{1}{6}$$

A olayı koşulken B olayının olasılığı,

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/6}{3/6} = \frac{1}{3}$$

B olayı koşulken A olayının olasılığı da,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{2/6} = \frac{1}{2}$$



# Bayes Teoremi

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Aynı şekilde  $A$  olayı verilmiş ise  $B$  olayının olasılığı şudur:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Bu iki denklem yeniden düzenlenip birbirlerine birleştirilirse,

$$P(A|B) P(B) = P(A \cap B) = P(B|A) P(A).$$

ifadesi bulunur. Bu **lemma** bazen olasılıklar için *çarpım kuralı* olarak anılmaktadır.

Her iki taraf da  $P(B)$  (eğer sıfır değilse) ile bölünürse, ortaya çıkan şu ifade Bayes teoremidir:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) P(A)}{P(B)}.$$



# Bayes Teoremi

- Örn: Okula giden her 36 çocuktan 1 tanesi Covid-19 hastalığına yakalanmaktadır ve test sonuçlarına göre, hastalıklı bir çocuğun testi %60 pozitif, sağlıklı bir çocuğun testi ise %20 pozitif sonuç vermektedir.

Bu bilgilere göre test sonucu pozitif olan bir çocuğun gerçekten hasta olma olasılığı nedir?

$P(A)$  : Çocuğun hasta olması olasılığı =  $1/36 = 0,027$


$P(B)$  : Testin pozitif çıkması olasılığı =  $1/36 * 0,6 + 35/36 * 0,2 = 0,211$

$P(A | B)$  : Pozitif çıkan testte hasta olma olasılığı

$P(B | A)$  : Hasta çocuğun testinin pozitif çıkma olasılığı =  $0,6$

$P(A | B) = P(B | A) * P(A) / P(B) = 0,6 * 0,027 / 0,211 = 0,077 \quad \%8$

# Naive Bayes Sınıflandırma

- Öznitelikler için iki temel varsayım
  - Bağımsız 
  - Eşit
- Sorumuza cevap özniteliği (olay) seçip hesaplıyoruz

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad X = (x_1, x_2, x_3, \dots, x_n)$$

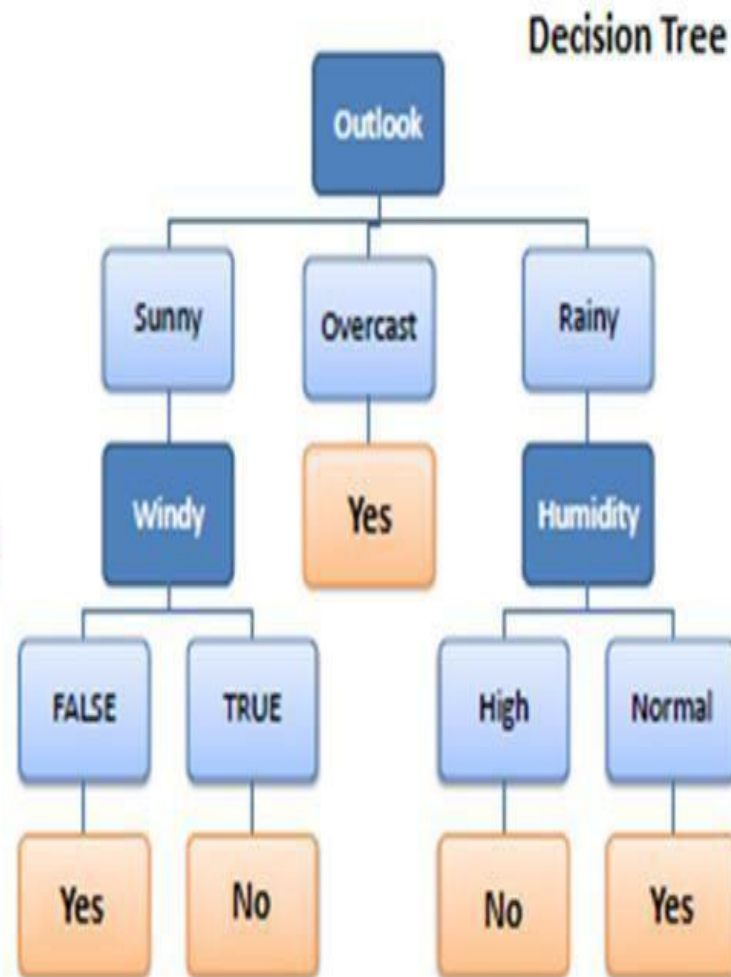
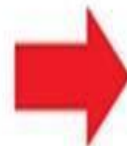
$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)}$$

$$P(y|x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i|y)}{P(x_1)P(x_2)\dots P(x_n)}$$

$$y = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i|y)$$

# GİNİ ENDEKSİ

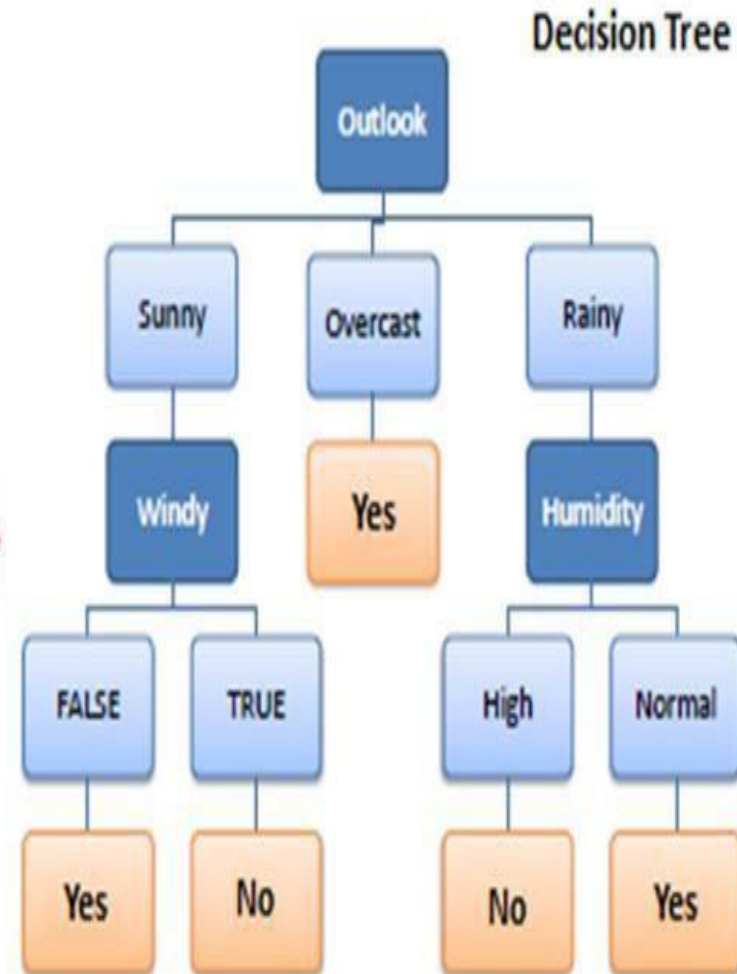
Predictors				Target
Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



# NAÏVE BAYES

today = (Sunny, Hot, Normal, False)

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



# NAÏVE BAYES

today = (Sunny, Hot, Normal, False)

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Outlook				
	Yes	No	P(Yes)	P(no)
Sunny	3	2	3/9	2/5
Overcast	4	0	4/9	0/5
Rainy	2	3	2/9	3/5
<b>Total</b>	9	5	100%	100%

Temperature				
	Yes	No	P(Yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
<b>Total</b>	9	5	100%	100%

Humidity				
	Yes	No	P(Yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
<b>Total</b>	9	5	100%	100%

Wind				
	Yes	No	P(Yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
<b>Total</b>	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
<b>Total</b>	14	100%



$$P(Yes|today) = \frac{P(Sunny|Outlook|Yes)P(Hot|Temperature|Yes)P(Normal|Humidity|Yes)P(No|Wind|Yes)P(Yes)}{P(today)}$$

$$P(No|today) = \frac{P(Sunny|Outlook|No)P(Hot|Temperature|No)P(Normal|Humidity|No)P(No|Wind|No)P(No)}{P(today)}$$

today = (Sunny, Hot, Normal, False)

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

	Yes	No	P(Yes)	P(no)
Sunny	3	2	3/9	2/5
Overcast	4	0	4/9	0/5
Rainy	2	3	2/9	3/5
<b>Total</b>	9	5	100%	100%

	Yes	No	P(Yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
<b>Total</b>	9	5	100%	100%

	Yes	No	P(Yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
<b>Total</b>	9	5	100%	100%

	Yes	No	P(Yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
<b>Total</b>	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
<b>Total</b>	14	100%

$$P(Yes|today) = \frac{P(Sunny|Outlook|Yes)P(Hot|Temperature|Yes)P(Normal|Humidity|Yes)P(No|Wind|Yes)P(Yes)}{P(today)}$$

$$P(No|today) = \frac{P(Sunny|Outlook|No)P(Hot|Temperature|No)P(Normal|Humidity|No)P(No|Wind|No)P(No)}{P(today)}$$

today = (Sunny, Hot, Normal, False)

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

	Yes	No	P(Yes)	P(no)
Sunny	3	2	3/9	2/5
Overcast	4	0	4/9	0/5
Rainy	2	3	2/9	3/5
Total	9	5	100%	100%

	Yes	No	P(Yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

	Yes	No	P(Yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

	Yes	No	P(Yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

$$P(Yes|today) \propto \frac{3}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} \approx 0,0212$$

$$P(No|today) \propto \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} \approx 0,0046$$



$$P(Yes|today) = \frac{P(Sunny|Outlook|Yes)P(Hot|Temperature|Yes)P(Normal|Humidity|Yes)P(No|Wind|Yes)P(Yes)}{P(today)}$$

$$P(No|today) = \frac{P(Sunny|Outlook|No)P(Hot|Temperature|No)P(Normal|Humidity|No)P(No|Wind|No)P(No)}{P(today)}$$

today = (Sunny, Hot, Normal, False)

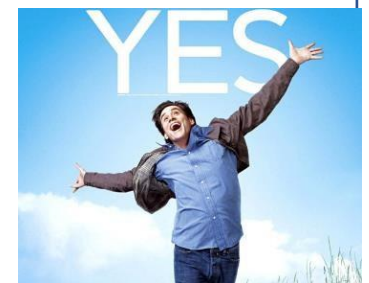
Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Outlook Temperature

$$P(Yes|today) = \frac{0,0212}{0,0212 + 0,0046} = 0,822368$$

$$P(No|today) = \frac{0,0046}{0,0212 + 0,0046} = 0,177632$$

$$P(Yes|today) > P(No|today)$$



$$P(Yes|today) \propto \frac{3}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} \approx 0,0212$$

$$P(No|today) \propto \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} \approx 0,0046$$