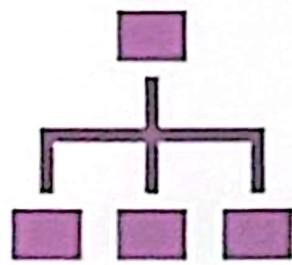


## SINIFLANDIRMA PERFORMANS DEĞERLENDİRME ÖLÇÜTLERİ

- Makine öğrenmesi modellerinin başarısını değerlendirme kriterleri

## GİRİŞ



- Sınıflandırma nedir?
- Performans değerlendirme neden önemlidir?

## TEMEL KAVRAMLAR

- Gerçek Pozitif (TP)

- Gerçek Negatif (TN)

- Yanlış Pozitif (FP)

- Yanlış Negatif (FN)

## TEMEL KAVRAMLAR

- Gerçek Pozitif (TP): Modelin doğru bir şekilde pozitif olarak tahmin ettiği örnekler.
- Gerçek Negatif (TN): Modelin doğru bir şekilde negatif olarak tahmin ettiği örnekler.
- Yanlış Pozitif (FP): Modelin yanlışlıkla pozitif olarak tahmin ettiği örnekler.
- Yanlış Negatif (FN): Modelin yanlışlıkla negatif olarak tahmin ettiği örnekler.

## HATA MATRİSİ - ÖRNEK

- Gerçek ve tahmin edilen değerlerin karşılaştırıldığı bir tablodur.
- İkili sınıflandırmada  $2 \times 2$  matris şeklinde gösterilir.

Gerçek \ Tahmin	Pozitif (Hasta)	Negatif (Sağlıklı)
Pozitif (Hasta)	$TP = 50$	$FN = 5$
Negatif (Sağlıklı)	$FP = 10$	$TN = 35$

Doğruluk  
(Accuracy)

Kesinlik  
(Precision)

Duyarlılık  
(Recall)

F1-Skoru

ROC Eğrisi  
ve AUC

## DEĞERLENDİRME ÖLÇÜTLERİ

---

# DOĞRULUK (ACCURACY)

- Tüm doğru tahminlerin toplam örneklerle oranıdır.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

**Örnek:** 100 hastadan 90'ı doğru tahmin edildiyse doğruluk =  $90/100 = \%90$

## KESİNLİK (PRECISION)

- Modelin pozitif tahminlerinin ne kadarının gerçekten pozitif olduğunu ölçer.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

**Örnek:** Kanser tespit eden bir modelde precision yüksek olmalıdır.

## DUYARLILIK (RECALL - SENSIVITY)

- Gerçek pozitif örneklerin ne kadarının doğru tahmin edildiğini gösterir.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

**Örnek:** Kanser testinde, yanlış negatifleri (FN) azaltmak için recall yüksek olmalıdır.

## FI-SKORU

- Precision ve Recall'un dengeli bir ölçüsüdür.

$$\text{FI-Skoru} = 2 * (\text{Kesinlik} * \text{Duyarlılık}) / (\text{Kesinlik} + \text{Duyarlılık})$$

**Örnek:** Dengesiz veri setlerinde FI-Skoru önemli bir metriktir.

# ROC EĞRİSİ

## AUC

- Receiver Operating Characteristic (ROC) eğrisi, modelin farklı eşik değerlerindeki performansını gösterir.
- AUC (Area Under Curve), modelin genel başarı oranını ölçer.

**Örnek:** AUC değeri 1'e ne kadar yakınsa, model o kadar iyidir.

## ROC EĞRİSİ VE AUC HESAPLAMA ADIMLARI

- 1) Modelin tahmin skorları belirlenir.
- 2) Farklı eşik değerleri için True Positive Rate (TPR) ve False Positive Rate (FPR) hesaplanır.
- 3) ROC eğrisi çizilir: X eksen FPR, Y eksen TPR'dır.
- 4) Eğri altındaki alan (AUC) hesaplanarak modelin başarısı değerlendirilir.

AUC = 1 mükemmel model, AUC = 0.5 rastgele model anlamına gelir.

## TRUE POSITIVE RATE (TPR) FALSE POSITIVE RATE (FPR)

TPR, modelin pozitif sınıfları doğru tahmin etme oranını gösterir.

$$TPR = \frac{TP}{TP + FN}$$

- TP (True Positive): Doğru şekilde pozitif olarak tahmin edilen örnekler.
- FN (False Negative): Gerçekte pozitif olup yanlışlıkla negatif tahmin edilen örnekler

FPR, gerçekte negatif olan örneklerin yanlışlıkla pozitif olarak tahmin edilme oranını gösterir.

$$FPR = \frac{FP}{FP + TN}$$

- FP (False Positive): Gerçekte negatif olup yanlışlıkla pozitif tahmin edilen örnekler.
- TN (True Negative): Doğru şekilde negatif olarak tahmin edilen örnekler

## ROC EĞRİSİNİN ŞEKLİ

### 1. Tek kırılma noktasına sahip ROC eğrileri:

- Eğer model sadece iki farklı olasılık değeri üretiyorsa (örneğin, yalnızca 0.3 ve 0.7), ROC eğrisinde tek bir kırılma noktası olur.
- Basit modellerde veya az sayıda farklı tahminin değerine sahip veri setlerinde görülebilir.

### 2. Basamak şeklinde ROC eğrileri:

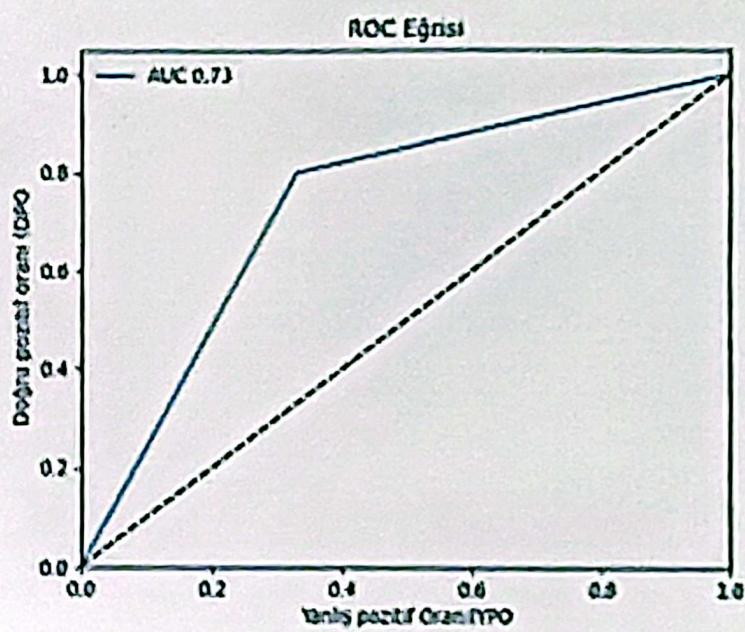
- Modelin tahminleri çok çeşitli değerler içerdilginde (örneğin, 0.1, 0.2, 0.3, ..., 0.9 gibi), ROC eğrisi daha kademeli olur.
- Küçük eşik değişimlerinde bile FPR ve TPR değişir, bu yüzden eğri daha fazla kırılma noktasına sahip olur.

Modelin tahmin çıktıları ne kadar farklı ve sürekli bir dağılıma sahipse, ROC eğrisi o kadar basamaklı olur.

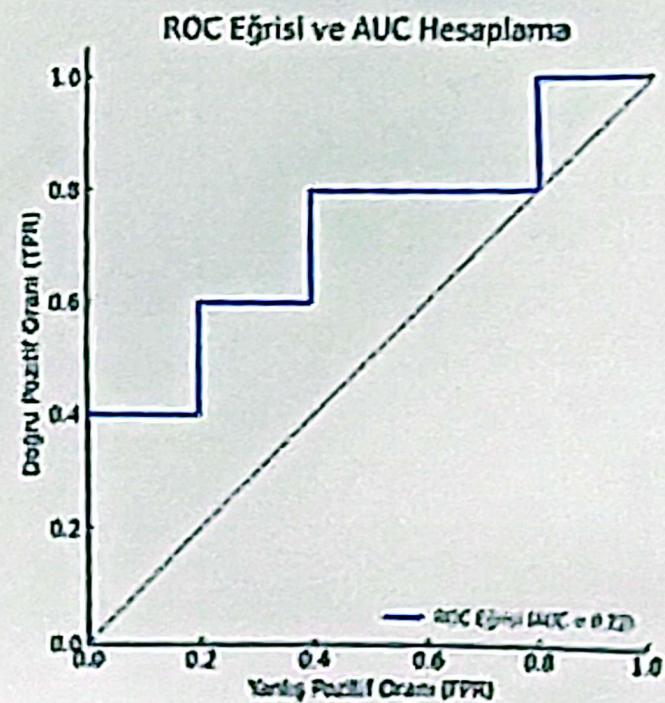
Eğer az sayıda farklı tahminin değeri varsa, ROC eğrisi daha az kırılmaya sahip olur.

# ROC EĞRİSİ ÖRNEK

ÖRNEK-I



ÖRNEK-II



## DEĞERLENDİRME ÖLÇÜTLERİ - ÖRNEK

Gerçek \ Tahmin

Pozitif (Hasta)

Negatif (Sağlıklı)

Pozitif (Hasta)

TP = 50

FN = 5

Negatif (Sağlıklı)

FP = 10

TN = 35

$$\text{Doğruluk (Accuracy)} = (TP + TN) / (TP + TN + FP + FN)$$

$$= (50 + 35) / (50 + 35 + 10 + 5)$$

$$= 85 / 100 = \%85$$

## DEĞERLENDİRME ÖLÇÜTLERİ - ÖRNEK

Gerçek \ Tahmin	Pozitif (Hasta)	Negatif (Sağlıklı)
Pozitif (Hasta)	TP = 50	FN = 5
Negatif (Sağlıklı)	FP = 10	TN = 35

$$\cdot \text{Kesinlik (Precision)} = \text{TP} / (\text{TP} + \text{FP})$$

$$= 50 / (50 + 10)$$

$$= 50 / 60 = \%83.3$$

## DEĞERLENDİRME ÖLÇÜTLERİ - ÖRNEK

Gerçek \ Tahmin

Pozitif (Hasta)

Negatif (Sağlıklı)

Pozitif (Hasta)

TP = 50

FP = 10

Negatif (Sağlıklı)

FN = 5

TN = 35

$$\cdot F1\text{-Skoru} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$= 2 * (0.833 * 0.909) / (0.833 + 0.909)$$

$$\approx \%86.9$$

## SONUÇ VE ÖNERİLER

- Modellere uygun değerlendirme metrikleri seçmek önemlidir.
- Farklı problemler için hangi ölçütün daha önemli olduğu değişebilir:  
**Tıbbi teşhis:** Yüksek duyarlılık (recall) önemlidir.  
**Spam filtreleme:** Yüksek kesinlik (precision) önemlidir.
- Modellere uygun değerlendirme metrikleri seçmek büyük önem taşır.
- Farklı metrikleri bir arada değerlendirerek daha sağlıklı sonuçlar alınabilir.