

# BLM 4088

Hafta4

# Sınıflandırma Problemlerindeki Metrikler

- İkili bir sınıflandırma durumunda, modelin performansını değerlendirmek için gerekli olan ana metrikler aşağıda verilmiştir.
- **Karışıklık matrisi** — Karışıklık matrisi, bir modelin performansını değerlendirirken daha eksiksiz bir sonuca sahip olmak için kullanılır. Aşağıdaki şekilde tanımlanmıştır:

		Tahmini sınıf	
		+	-
Gerçek sınıf	+	<b>TP</b> True Positives	<b>FN</b> False Negatives Type II error
	-	<b>FP</b> False Positives Type I error	<b>TN</b> True Negatives

# Sınıflandırma Problemlerindeki Metrikler

- TP: Lastikçinin patlak olan lastiğine *“abi bu patlak”* demesi (Malumun ilamı, bravo!)
- TN: Doktorun hasta olan birine *“bişeyin yok turp gibisin”* demesi
- FP: Üstünüzde t-shirt ve kot ile AVM güvenliğinden geçerken cebinizdeki anahtarın ötmesi ve güvenlikçinin size kötü adam muamelesi çekmesi, *“kenara geçip ceplerinizi boşaltın”* demesi (cebinizde tehlikeli bir şey var!!!)
- FN: Komşunun camını kırdıktan sonra annenizin *“benim oğlum yapmaz öyle bişey o kırmadı”* diye komşuya kızması

# Sınıflandırma Problemlerindeki Metrikler

- *TP (True positive — Doğru Pozitif)*: Hastaya hasta demek.
- *FP (False positive — Yanlış Pozitif)*: Hasta olmayana hasta demek.
- *TN (True negative — Doğru Negatif)*: Hasta olmayana hasta değil demek.
- *FN (False negative — Yanlış Negatif)*: Hasta olana hasta değil demek.

# Sınıflandırma Problemlerindeki Metrikler

- **True Positive (TP):** Modelimiz (Çoban) “Kurt var!” der ve gerçekte de kurt vardır. Modelimizin ***doğru*** tahmin ettiği ihtimallerden biridir.
- **True Negative (TN):** Modelimiz (Çoban) “Kurt yok” der ve kurt yoktur. Modelimizin ***doğru*** tahmin ettiği ihtimallerden biridir.
- **False Positive (FP):** Modelimiz (Çoban) “Kurt var!” der ama kurt yoktur. Modelimizin ***yanlış*** tahmin ettiği ihtimallerden biridir.
- **False Negative (FN):** Modelimiz (Çoban) “Kurt yok” der ama kurt vardır. Modelimizin ***yanlış*** tahmin ettiği ihtimallerden biridir.

# Tahmin

## Gerçek

YES

NO

YES

### True Positive

- Kurt vardır.
- Çoban "Kurt geldi!" der.
- Çoban kahraman ilan edilir.

TP sayısı: 1

### False Positive

- Kurt yoktur.
- Çoban "Kurt geldi!" der.
- Köylüler çobana kızar.

FP sayısı: 1

NO

### False Negative

- Kurt vardır.
- Çoban "Kurt yok" der.
- Kurt kuzuları yer.

FN sayısı: 8

### True Negative

- Kurt yoktur.
- Çoban "Kurt yok" der.
- Herkes iyi.

TN sayısı: 90

# Sınıflandırma metrikleri

- **Ana metrikler** — Sınıflandırma modellerinin performansını değerlendirmek için aşağıda verilen metrikler yaygın olarak kullanılmaktadır:

Metrik	Formül	Açıklama
Doğruluk	$\frac{TP + TN}{TP + TN + FP + FN}$	Modelin genel performansı
Kesinlik	$\frac{TP}{TP + FP}$	Doğru tahminlerin ne kadar kesin olduğu
Geri çağırma	$\frac{TP}{TP + FN}$	Gerçek pozitif örneklerin oranı
Specificity	$\frac{TN}{TN + FP}$	Gerçek negatif örneklerin oranı
F1 skoru	$\frac{2TP}{2TP + FP + FN}$	Dengesiz sınıflar için yararlı hibrit metrik

# Sınıflandırma metrikleri

- ***Accuracy = Doğru Tahminlerin Sayısı / Tüm Tahminlerin Sayısı***
- Accuracy değerini dikkate almak her sınıfta eşit sayıda gözlem varsa mantıklı olabilir. Bunun dışında oldukça yanıltıcıdır. Örneğin yukarıdaki modelin accuracy değerini hesaplayalım.
- ***Accuracy = (1+90) / (1+90+1+8) = 0.91***
- Bu değer gayet güzel gibi görünebilir ama sayılara tekrar bakıldığında kurdun 9 kez geldiğini ve sadece 1 kez doğru tahmin ediliyor. Bu çok kötü bir sonuç. Halbuki Accuracy skorumuz %91'di.



# Sınıflandırma metrikleri

- *Precision = Pozitif olarak tahmin edilenlerin gerçekte kaçta kaçısı doğru.*
- *Precision =  $TP / (TP+FP)$*
- *Precision =  $1 / (1+1) = 0.5$*
- *Recall = Model, pozitif classların kaçta kaçını yakalayabiliyor.*
- *Recall =  $TP / (TP+FN)$*
- *Recall =  $1 / (1 + 8) = 0.11$*

# Sınıflandırma metrikleri

- Bu metriklerin önemi, modeli hangi amaçla geliştirdiğinize göre değişim göstermektedir. Örneğin evinizi korumak için, hırsızları yakalamak üzere eğittiğiniz modelde **Recall** değeri çok önemlidir. **Recall** değeri oldukça yüksek olmalıdır. Tam tersine, diğer hayvan fotoğrafları arasından kedi fotoğraflarını ayırabilen modelde, daha fazla kedi fotoğrafını yakalamak için **Recall** değeri esnetilebilir.

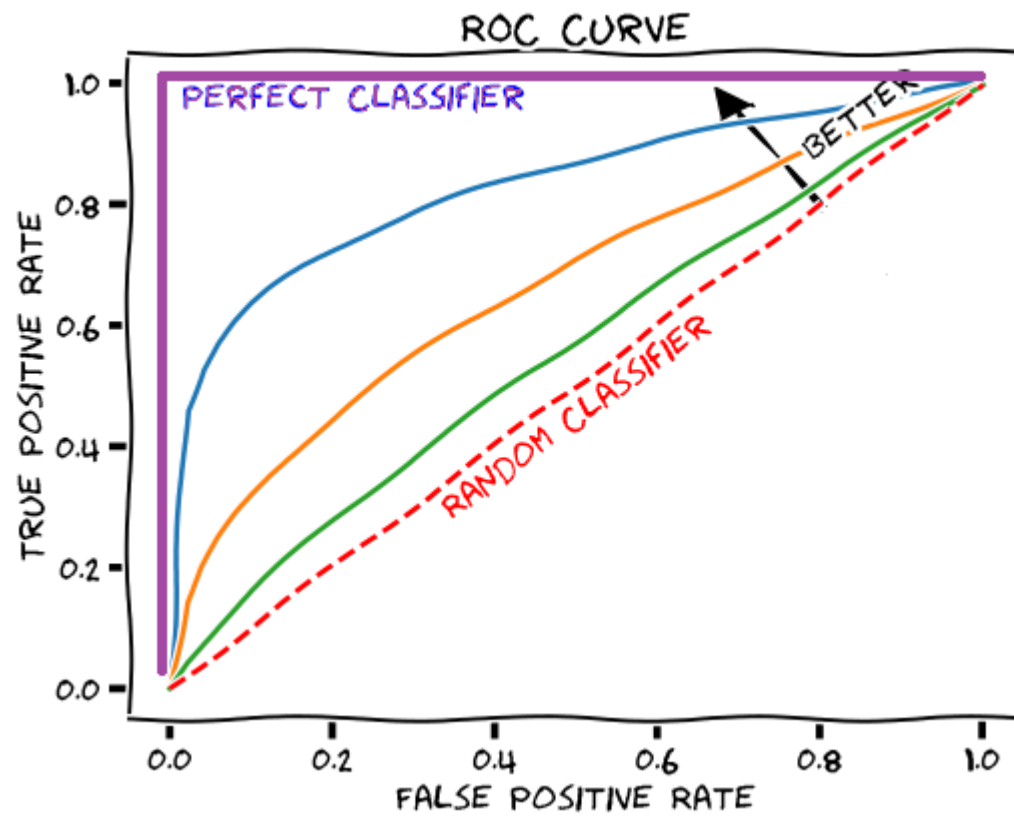
# Sınıflandırma metrikleri

- **F1 Score**
- F1 Score, Precision ve Recall değerlerinin ağırlıklı (harmonik) ortalamasıdır.
- **$F1\ Score = 2 * (Recall * Precision) / (Recall + Precision)$**
- F1 Score, özellikle eşit olmayan bir sınıf dağılımınız varsa, Accuracy değerinden daha kullanışlı olabilir.
- **$F1\ Score = 2 * (0.11 * 0.5) / (0.11 + 0.5) = 0,18$**

# Sınıflandırma metrikleri

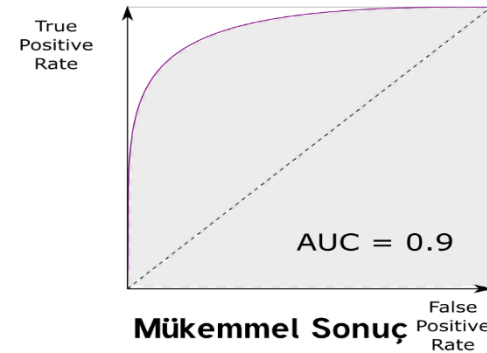
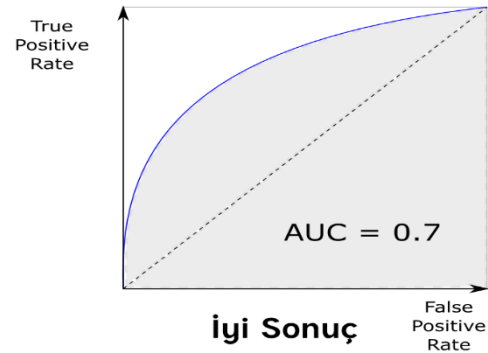
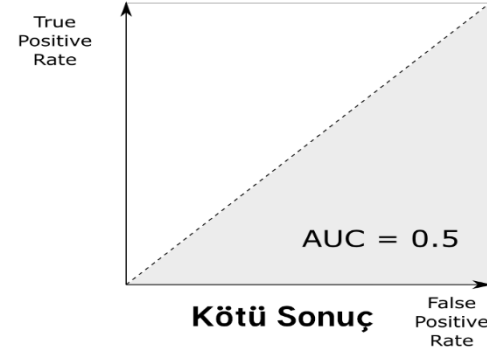
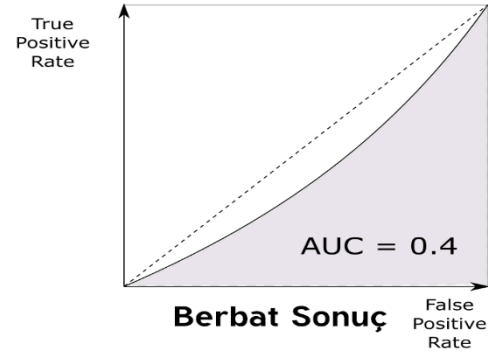
- ROC ve AUC
- Sınıflandırma problemlerinde olasılık değerini sınıflandırmak için eşik değere (threshold) ihtiyaç duyulur. Örneğin 0.5 gibi; Sonuçları, 0.5'in altı ve 0.5'in üstü olacak şekilde böler. **ROC Curve**, eşik değerinin performansını gösteren eğridir. Tablonun x ekseninde FP Rate, y ekseninde ise TP Rate (Recall) vardır.
- **$TPR = TP / (TP + FN)$  (Sensitivity olarak da biliniyor.)**
- **$FPR = FP / (FP + TN)$  (1-Specificity olarak da biliniyor.)**

# ROC ve AUC



# ROC ve AUC

- ROC eğrisinin altında kalan yere AUC (Area Under the ROC Curve) denir. Positive Class'ı Negative Class'tan ne kadar ayırabildiğini verir. Alan arttıkça ayırım yeteneği artar.



# Kaynak

- <https://developers.google.com/machine-learning/crash-course/classification/accuracy>
- <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>
- <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>