

Başarım Metrikleri

- Problemlerin çözümü genellikle pek çok farklı algoritmayı farklı parametrelerle yeniden denemeyi içerir.
- Bunların arasından en iyisini nasıl seçebiliriz?
- **Neye göre “en iyi”?**

Accuracy (kesinlik):

a = Doğru tahmin edilen test örneği sayısı

b = Yanlış tahmin edilen test örneği sayısı

$$accuracy = \frac{a}{a + b}$$

Örnek

- 50 soruluk bir sınav
- 35 doğru cevap, 15 yanlış cevap
- $accuracy = 35/50 = 0.7$

- **False Positives (Hatalı Pozitif)**
 - Sistem **DOĞRU** tahmin etti ama **YANLIŞ** olmalıydı.
 - Yanlış alarmlar.
 - **Type I Error (istatistik)**
 - **Örnek:** Çoban köylülere kurt geliyor dedi, ama kurt gelmedi.
- **False Negatives (Hatalı Negatif)**
 - Sistem **YANLIŞ** tahmin etti ama **DOĞRU** olmalıydı.
 - Kaçırılan örnekler.
 - **Type II Error (istatistik)**
 - **Örnek:** Çoban köylülere kurt gelmiyor dedi, ama kurt geldi.
- **True Positives (Doğru Pozitif), True Negatives (Doğru Negatif)**

Confusion Matrix (Karışıklık Matrisi)

		actual	
		+	-
classifier	+	TP good!	FP bad! (Type I error)
	-	FN bad! (Type II error)	TN good!

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	21	6
	Class 2	7	41

TP =

TN =

FP =

FN =

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	21	6
	Class 2	7	41

TP = 21

TN =

FP =

FN =

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	21	6
	Class 2	7	41

TP = 21

TN = 41

FP =

FN =

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	21	6
	Class 2	7	41

TP = 21

TN = 41

FP = 6

FN =

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	21	6
	Class 2	7	41

TP = 21

TN = 41

FP = 6

FN = 7

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	21	6
	Class 2	7	41

TP = 21

TN = 41

FP = 6

FN = 7

$$accuracy = \frac{21 + 41}{21 + 6 + 7 + 41} = \frac{62}{75}$$

Confusion Matrix (Karışıklık Matrisi)

		Gerçek Sınıf	
		Class 1 (POS)	Class 2 (NEG)
Tahmin Edilen Sınıf	Class 1 (POS)	21 (TP)	6 (FP)
	Class 2 (NEG)	7 (FN)	41 (TN)

TP = 21

TN = 41

FP = 6

FN = 7

$$accuracy = \frac{21 + 41}{21 + 6 + 7 + 41} = \frac{62}{75}$$

Accuracy (Kesinlik) Problemleri

- **Bir binary-classification problemi düşünün:**
 - class 1 test örneği sayısı = 10
 - class 2 test örneği sayısı = 9990
- **Bütün örnekleri class 2 tahmin edersek accuracy değeri kaç olur?**

Accuracy (Kesinlik) Problemleri

- **Bir binary-classification problemi düşünün:**
 - class 1 test örneği sayısı = 10
 - class 2 test örneği sayısı = 9990
- **Bütün örnekleri class 2 tahmin edersek accuracy değeri kaç olur?**
 - $9990 / 10000 = 99.9\%$
 - Model başarımı çok yüksek gibi duruyor, fakat aslında hiçbir şey öğrenmiyor.
 - Class 1 **asla** doğru tahmin edilmeyecek.
 - Böyle bir durumda accuracy değeri yanlış yönlendirmeye müsait.
 - Örnek durum?

Accuracy (Kesinlik) Problemleri

- **Bir binary-classification problemi düşünün:**
 - class 1 test örneği sayısı = 10
 - class 2 test örneği sayısı = 9990
- **Bütün örnekleri class 2 tahmin edersek accuracy değeri kaç olur?**
 - $9990 / 10000 = 99.9\%$
 - Model başarıımı çok yüksek gibi duruyor, fakat aslında hiçbir şey öğrenmiyor.
 - Class 1 **asla** doğru tahmin edilmeyecek.
 - Böyle bir durumda accuracy değeri yanlış yönlendirmeye müsait.
 - Örnek durum?
- **Kanser hastalığı tespiti. Kişi geçmişine dayanarak kimin kanser olabileceğini tahmin edin. class 1 = kanser, class 2 = sağlıklı.**

Accuracy (Kesinlik) Problemleri

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	0	0
	Class 2	10	9990

TP = 0

TN = 9990

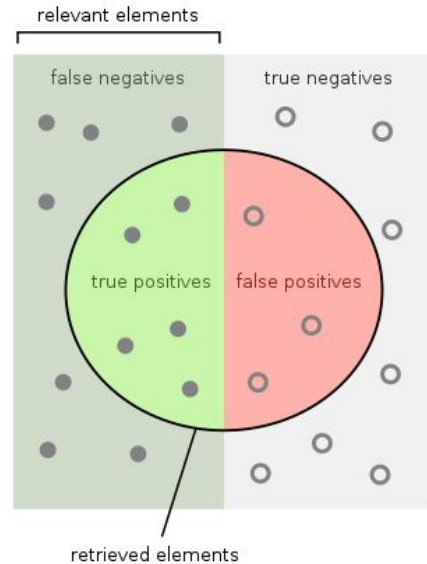
FP = 0

FN = 10

$$accuracy = \frac{0 + 9990}{0 + 0 + 9990 + 10} = 99.9\%$$

Precision, Recall, F-measure

- **Precision (hassaslık):** TP değerinin bütün pozitif tahminlere oranı
- **Recall (Sensitivity, TPR, duyarlılık):** TP değerinin bütün pozitif örneklere oranı
- **Specificity (TNR):** TN değerinin bütün negatif örneklere oranı
- **F-skoru:** Precision ve Recall değerlerinin harmonik ortalaması.



How many retrieved items are relevant?

Precision =



How many relevant items are retrieved?

Recall =



$$P = \frac{TP}{TP + FP}$$

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

$$F = \frac{2PR}{P + R}$$

Örnekler

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	1	0
	Class 2	9	90

P =

R =

F₁ =

$$P = \frac{TP}{TP + FP}$$

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

$$F = \frac{2PR}{P + R}$$

Örnekler

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	1	0
	Class 2	9	90

$$P = 1$$

$$R = 0.1$$

$$F_1 = 0.18$$

$$P = \frac{TP}{TP + FP}$$

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

$$F = \frac{2PR}{P + R}$$

Örnekler

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	10	50
	Class 2	0	40

P =

R =

F₁ =

$$P = \frac{TP}{TP + FP}$$

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

$$F = \frac{2PR}{P + R}$$

Örnekler

		Gerçek Sınıf	
		Class 1	Class 2
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	10	50
	Class 2	0	40

$$P = 1/6$$

$$R = 1$$

$$F_1 = 0.29$$

$$P = \frac{TP}{TP + FP}$$

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

$$F = \frac{2PR}{P + R}$$

Örnekler

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	9	1
	Class 2	1	89

P =

R =

F₁ =

$$P = \frac{TP}{TP + FP}$$

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

$$F = \frac{2PR}{P + R}$$

Örnekler

		Gerçek Sınıf	
		Class 1	Class 2
Tahmin Edilen Sınıf	Class 1	9	1
	Class 2	1	89

$$P = 0.9$$

$$R = 0.9$$

$$F_1 = 0.9$$

$$P = \frac{TP}{TP + FP}$$

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

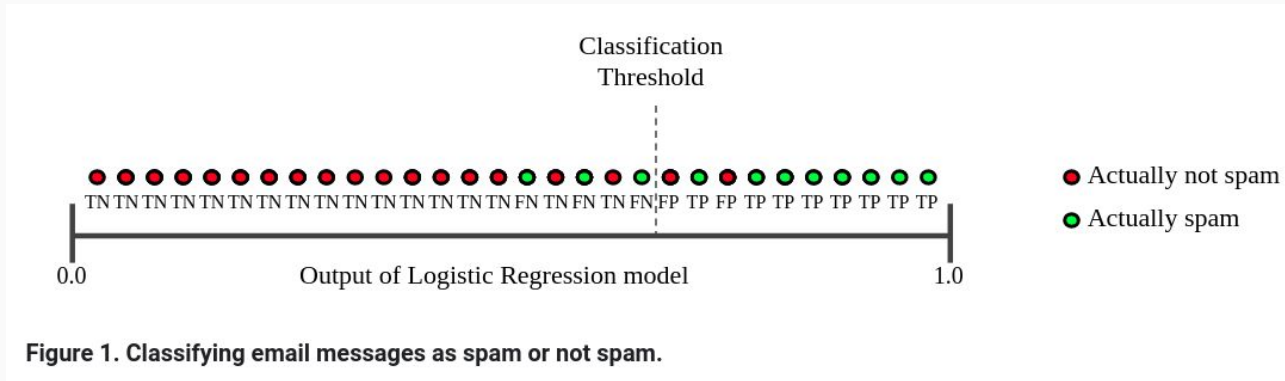
$$F = \frac{2PR}{P + R}$$

Thresholding (Eşikleme)

- **Bir binary-classification problemi düşünün.**
 - Modeliniz (0,1) demek yerine bir olasılık dönüyor olsun.
 - **Örnek 1:** Yarın hava yağmurlu mu değil mi tahmin edilmeye çalışılıyor. Model yarın havanın yağışlı olma ihtimalinin 51% olduğunu söylüyor.
 - **Örnek 2:** Gelen bir mail spam mı değil mi tespit edilmeye çalışılıyor. Model bir mailin spam olma ihtimalinin 78% olduğunu söylüyor.
 - **Sınıflandırma neye göre yapılabilir?**

Thresholding (Eşikleme)

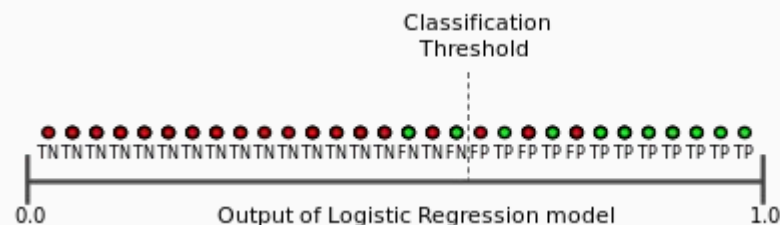
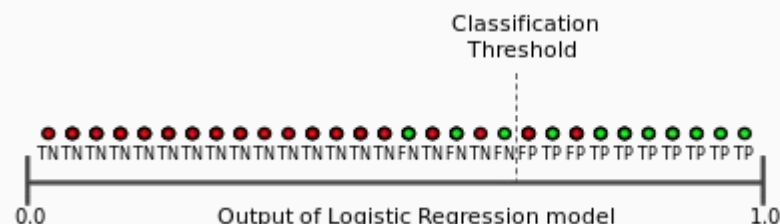
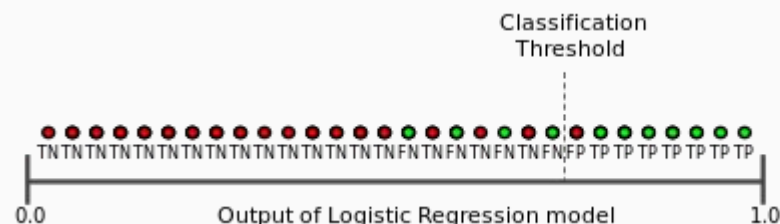
- **Bir binary-classification problemi düşünün.**
 - Modeliniz (0,1) demek yerine bir olasılık dönüyor olsun.
 - **Örnek 1:** Yarın hava yağmurlu mu değil mi tahmin edilmeye çalışılıyor. Model yarın havanın yağışlı olma ihtimalinin 51% olduğunu söylüyor.
 - **Örnek 2:** Gelen bir mail spam mı değil mi tespit edilmeye çalışılıyor. Model bir mailin spam olma ihtimalinin 78% olduğunu söylüyor.
 - **Sınıflandırma neye göre yapılabilir?**



- Uygun bir threshold nasıl belirleriz? <http://www.navan.name/roc/>

Thresholding (Eşikleme)

Toplam 30 test örneği

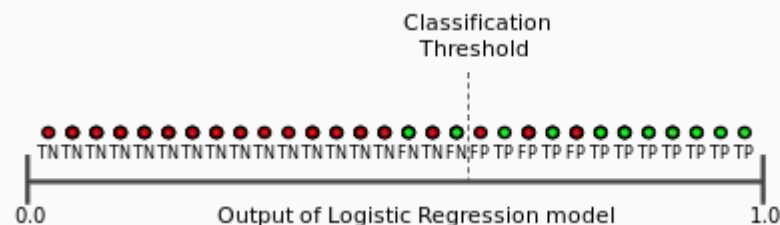
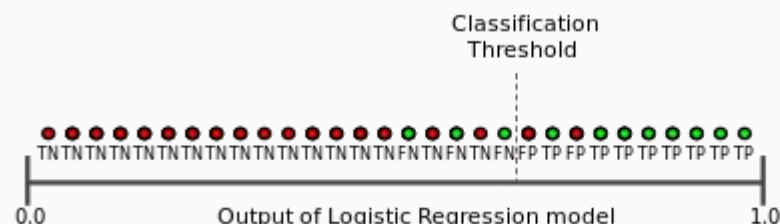
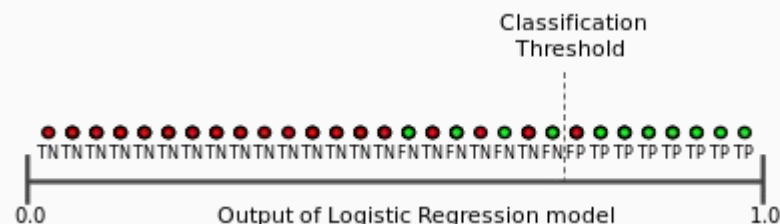


P:
R:

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam		
	Not Spam		

Thresholding (Eşikleme)

Toplam 30 test örneği

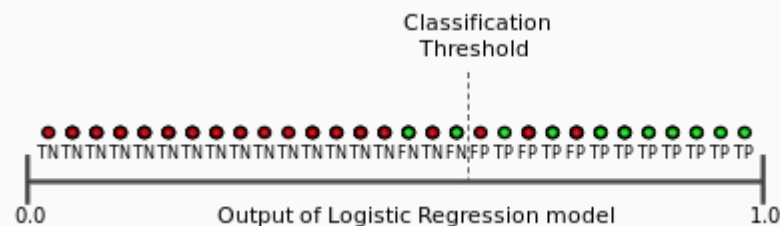
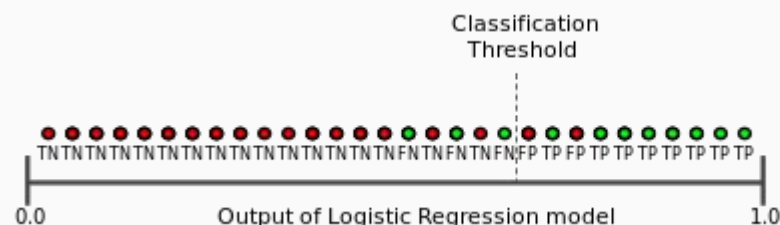
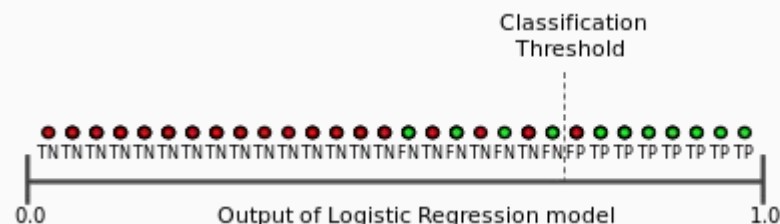


P:
R:

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

Thresholding (Eşikleme)

Toplam 30 test örneği

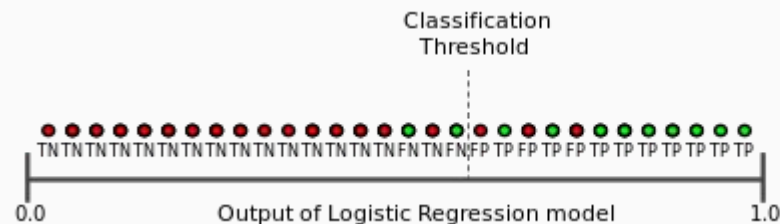
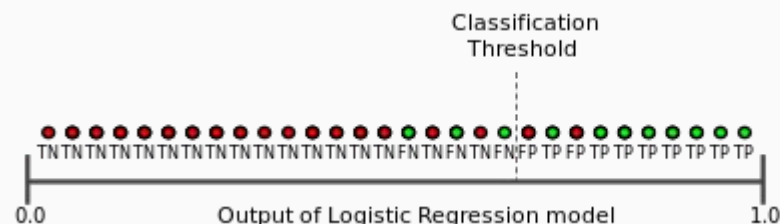
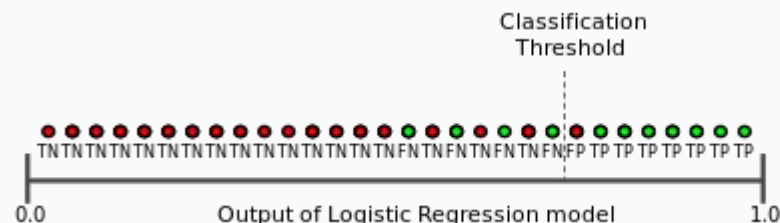


P: 0.88
R: 0.64

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

Thresholding (Eşikleme)

Toplam 30 test örneği



P: 0.88
R: 0.64

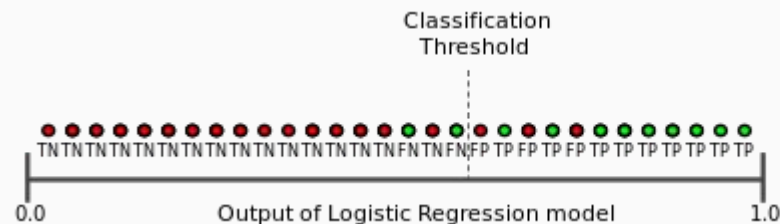
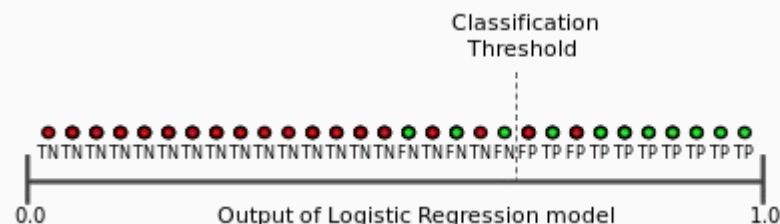
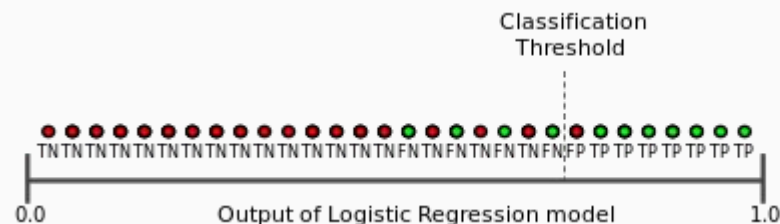
		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

P:
R:

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam		
	Not Spam		

Thresholding (Eşikleme)

Toplam 30 test örneği



P: 0.88
R: 0.64

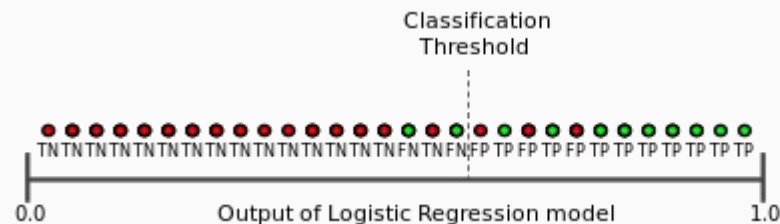
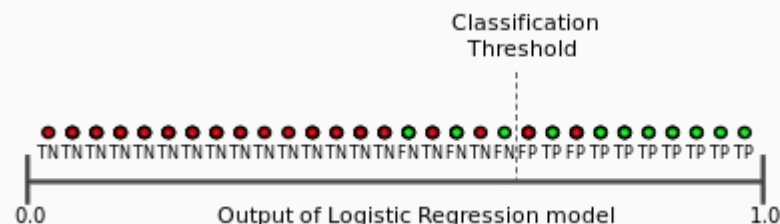
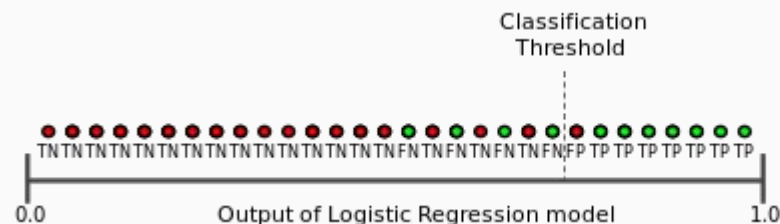
		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

P:
R:

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	8	2
	Not Spam	3	17

Thresholding (Eşikleme)

Toplam 30 test örneği



P: 0.88
R: 0.64

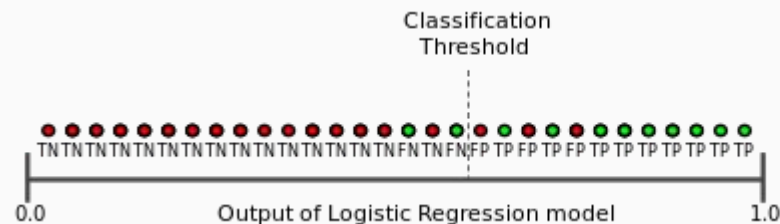
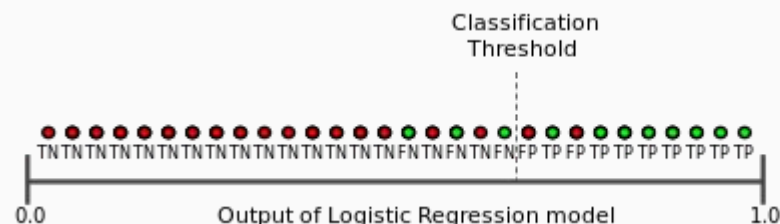
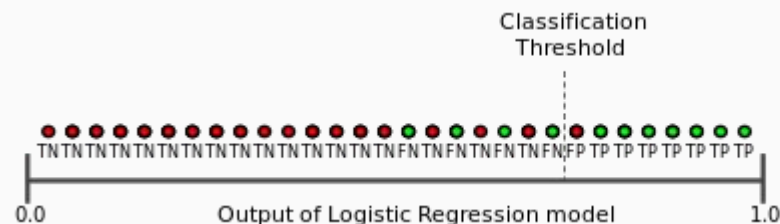
		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

P: 0.80
R: 0.73

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	8	2
	Not Spam	3	17

Thresholding (Eşikleme)

Toplam 30 test örneği



P: 0.88
R: 0.64

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

P: 0.80
R: 0.73

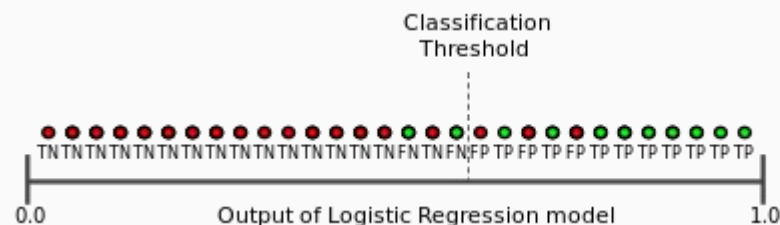
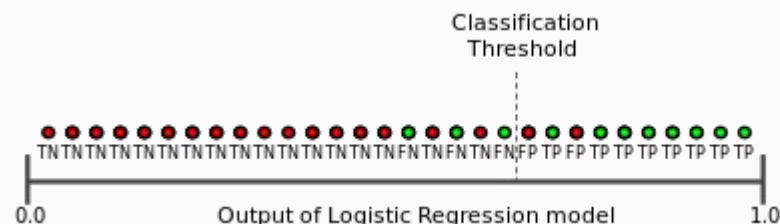
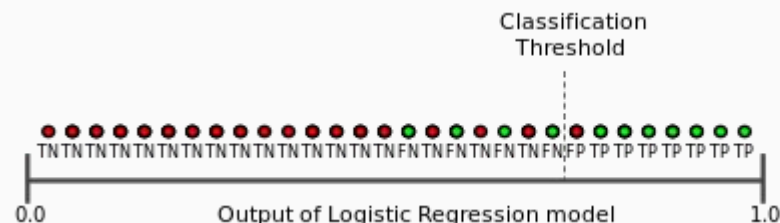
		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	8	2
	Not Spam	3	17

P:
R:

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam		
	Not Spam		

Thresholding (Eşikleme)

Toplam 30 test örneği



P: 0.88
R: 0.64

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

P: 0.80
R: 0.73

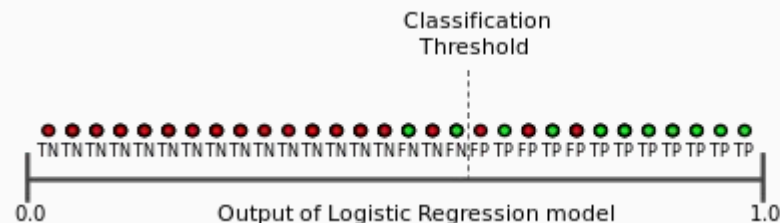
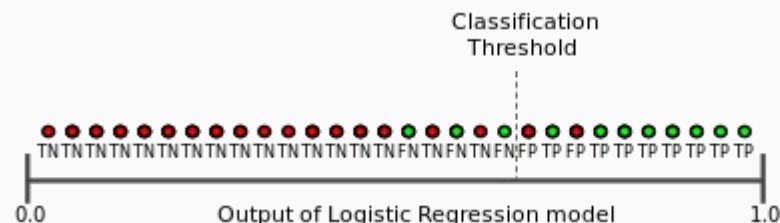
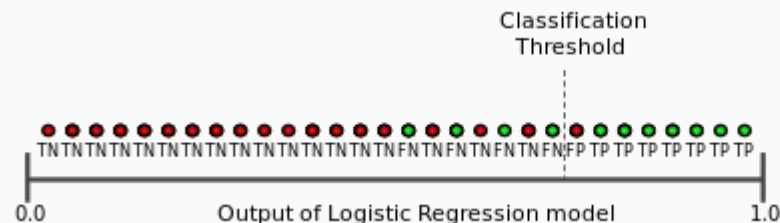
		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	8	2
	Not Spam	3	17

P:
R:

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	9	3
	Not Spam	2	16

Thresholding (Eşikleme)

Toplam 30 test örneği



P: 0.88
R: 0.64

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	7	1
	Not Spam	4	18

P: 0.80
R: 0.73

		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	8	2
	Not Spam	3	17

P: 0.75
R: 0.82

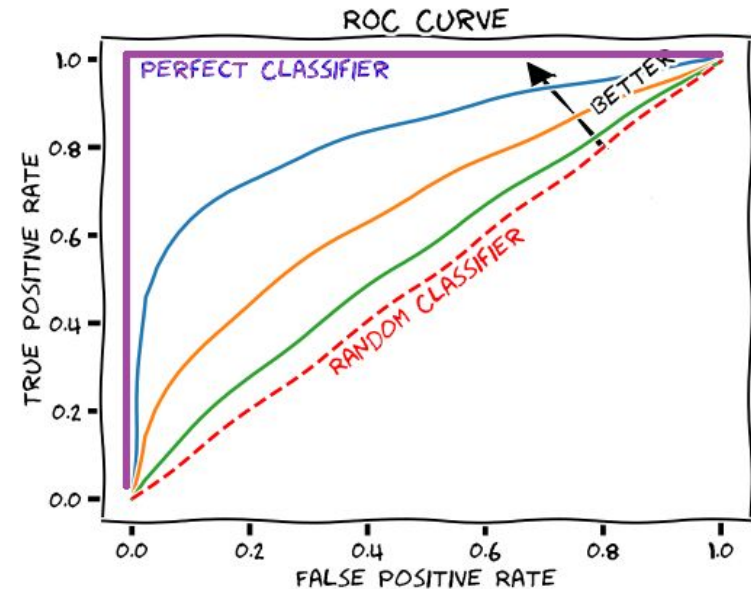
		Gerçek Sınıf	
		Spam	Not Spam
Tahmin Edilen Sınıf	Spam	9	3
	Not Spam	2	16

ROC Curve: receiver operating characteristic curve (alıcı işletim karakteristiği eğrisi)

- Bir sınıflandırma modelinin bütün sınır değerlerindeki performansını gösteren eğriye denir. TPR ve FPR değerleri kullanılarak oluşturulur.
- **TPR (True Positive Rate):** $TP / (TP + FN)$
- **TNR (True Negative Rate):** $TN / (FP + TN)$

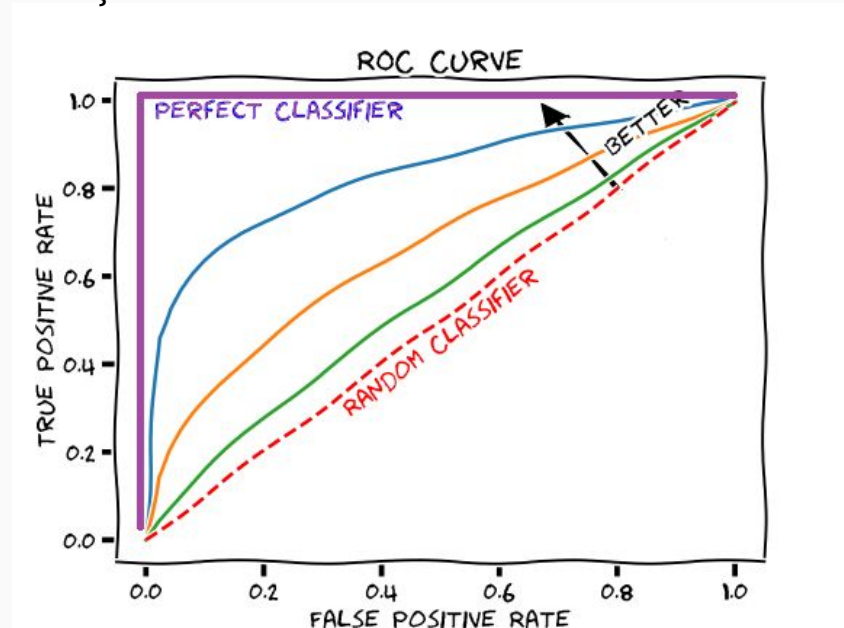
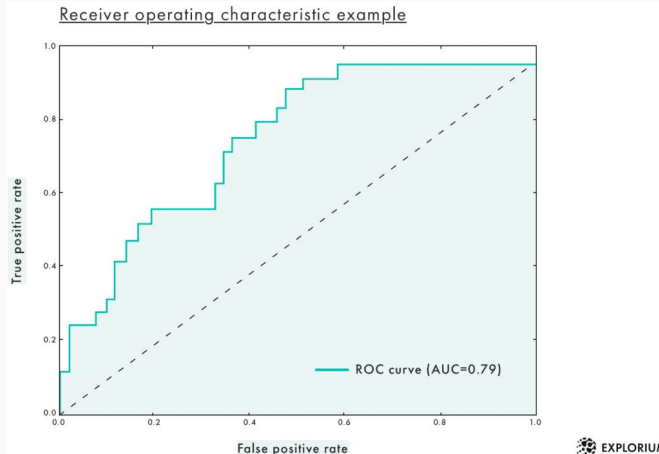
ROC Curve: receiver operating characteristic curve (alıcı işletim karakteristiği eğrisi)

- Bir sınıflandırma modelinin bütün sınır değerlerindeki performansını gösteren eğriye denir. TPR ve FPR değerleri kullanılarak oluşturulur.
- **TPR (True Positive Rate):** $TP / (TP + FN)$
- **TNR (True Negative Rate):** $TN / (FP + TN)$



ROC Curve: receiver operating characteristic curve (alıcı işletim karakteristiği eğrisi)

- Bir sınıflandırma modelinin bütün sınır değerlerindeki performansını gösteren eğriye denir. TPR ve FPR değerleri kullanılarak oluşturulur.
- **TPR (True Positive Rate):** $TP / (TP + FN)$
- **TNR (True Negative Rate):** $TN / (FP + TN)$



PR Curve: Precision-Recall Curve

- Farklı eşik değerleri için P-R ilişkisini görselleştiren eğridir.
- **AUC_{pr}**: PR Curve altında kalan alanı ifade eder.

PR Curve imbalanced datasetlerde azınlık pozitif için daha hassastır.

